

## KOMPARASI REGRESI ZIP DAN REGRESI ZINB PADA DATA KEMATIAN BAYI DI JAWA BARAT

DINA JUMIATUL FITRI<sup>1</sup>, ANGGRAINY TOGI MARITO SIREGAR<sup>2</sup>, HAPPY ALYZHYA HAA<sup>3</sup>, RASI KASIM SAMOSIR<sup>4</sup>

Program Studi Statistika, FMIPA, Universitas Cenderawasih, Jayapura<sup>1</sup>, Program Studi Matematika, FMIPA, Universitas Cenderawasih, Jayapura<sup>2,3</sup>, Program Studi Planologi, FT, Universitas Cenderawasih, Jayapura<sup>4</sup>  
Email: dinajumiatulfitri@gmail.com<sup>1</sup>

**Abstract:** *Poisson regression analysis is a nonlinear regression that is usually used for discrete data and assumes equidispersion. In practice, there is often a violation of assumptions equidispersion, one of the violations is overdispersion (the variance is greater than the average value). One of the causes of overdispersion is the excessive number of zero values (Excess Zero) on the response variable. Excess zeros can be seen in the proportion of response variables which is zero greater than any other discrete data. There are many methods to overcome overdispersion: Zero Inflated Poisson (ZIP) regression and Zero Inflated regression Negative Binomial (ZINB). The purpose of this study is to determine the regression model which is better used on data that experience overdispersion. Data used to analyze ZIP and ZINB regression is the data of infant death in West Java Province in 2021. Based on the study's result, it is known that the Akaike Information Criterion (AIC) value in the ZINB regression is smaller than the ZIP regression AIC value. So the ZINB regression is better used as well as the factor of infant death.*

**Keywords:** *Overdispersion; Poisson Regression; ZIP; ZINB; Infant Death.*

**Abstrak:** Analisis regresi Poisson adalah regresi nonlinier yang biasanya digunakan untuk data diskrit dan mengasumsikan equidispersi. Dalam praktiknya, sering terjadi pelanggaran asumsi equidispersi, salah satunya adalah overdispersi (varians lebih besar daripada nilai rata-rata). Salah satu penyebab overdispersi adalah jumlah nilai nol yang berlebihan (Excess Zero) pada variabel respon. Kelebihan nilai nol dapat dilihat pada proporsi variabel respon yang nolnya lebih besar daripada data diskrit lainnya. Ada banyak metode untuk mengatasi overdispersi: regresi Zero Inflated Poisson (ZIP) dan regresi Zero Inflated Negative Binomial (ZINB). Tujuan penelitian ini adalah untuk menentukan model regresi mana yang lebih baik digunakan pada data yang mengalami overdispersi. Data yang digunakan untuk menganalisis regresi ZIP dan ZINB adalah data kematian bayi di Provinsi Jawa Barat tahun 2021. Berdasarkan hasil penelitian, diketahui bahwa nilai Akaike Information Criterion (AIC) pada regresi ZINB lebih kecil daripada nilai AIC regresi ZIP. Oleh karena itu, regresi ZINB lebih baik digunakan sebagai faktor penyebab kematian bayi.

**Kata Kunci:** Overdispersi; Regresi Poisson; ZIP; ZINB; Kematian Bayi.

### A. Pendahuluan

Analisis regresi merupakan analisis statistika yang digunakan untuk memodelkan hubungan antara variabel respon dengan satu atau lebih variabel prediktor. Analisis regresi umumnya digunakan untuk menganalisis data variabel respon yang berupa data kontinu. Namun dalam beberapa aplikasinya, variabel respon yang akan dianalisa dapat berupa data diskrit. Salah satu model regresi yang dapat digunakan untuk menganalisis hubungan antara

variabel respon berupa data diskrit dan variabel prediktor berupa data kontinu, diskrit atau campuran adalah model regresi Poisson.

Analisis regresi Poisson digunakan untuk menganalisis hubungan antara sebuah variabel dependen yang menyatakan data terhitung atau data count berdistribusi Poisson dengan satu atau lebih variabel independen. Data count yang dimaksud misalnya adalah banyaknya kejadian dalam interval waktu, ruang, atau volume tertentu. Dalam analisis regresi Poisson, variabel independen yang digunakan dapat berjenis kontinu. Pada model regresi Poisson terdapat beberapa asumsi yang harus dipenuhi diantaranya adalah variabel respon berdistribusi Poisson, tidak terdapat multikolinearitas antar variabel prediktor dan *equidispersi* yaitu  $E(y) = Var(y) = \mu$ . Namun dalam analisis data diskrit dengan menggunakan model regresi Poisson terkadang terjadi pelanggaran asumsi tersebut, dimana nilai variansinya lebih besar dari nilai mean (*overdispersi*) atau varian lebih kecil dari nilai mean (*underdispersi*). Untuk menangani masalah *overdispersi*, diperlukan model-model statistik diantaranya adalah model regresi *Zero-Inflated Poisson* (ZIP) dan model regresi *Zero-Inflated Negative Binomial* (ZINB). *Overdispersi* data count dapat terjadi jika mengalami *excess zero*, yaitu kondisi dimana proporsi nilai nol pada data variabel respon lebih besar dari nilai lainnya. Pada kondisi *excess zero*, regresi Poisson pada dasarnya tetap dapat diestimasi menggunakan model regresi *Zero Inflated Poisson* (ZIP). Akan tetapi pada kondisi *excess zero* dan *overdispersi*, model regresi *Zero Inflated Poisson* (ZIP) tidak sesuai untuk memodelkan data dan model yang akan terbentuk menghasilkan estimasi yang bias. Hal ini menyebabkan adanya pengembangan model-model statistik untuk mengatasi masalah *excess zero* dan *overdispersi* yaitu dengan menggunakan model regresi *Zero Inflated Negative Binomial* (ZINB).

Salah satu contoh dimana variabel responnya diskrit dan banyaknya kejadian yang jarang terjadi (*rare event*) kasus kematian bayi. Kematian bayi didefinisikan sebagai jumlah kematian bayi usia 0-11 bulan (termasuk neonatal/bayi usia 0-28 hari) yang meninggal di suatu wilayah pada kurun waktu tertentu sedangkan Angka Kematian Bayi (AKB) adalah jumlah bayi usia 0-11 bulan (termasuk neonatal/baru lahir) yang meninggal di suatu wilayah pada kurun waktu tertentu dibagi 1000 kelahiran hidup di wilayah tersebut dan pada kurun waktu yang sama. Adapun target yang ditetapkan adalah menurunkan AKB menjadi 16 per 1000 kelahiran hidup pada tahun 2024.

Angka Kematian Bayi (AKB) Indonesia tahun 2020 masih menempati posisi lima tertinggi di Asia Tenggara yaitu sekitar 20 per 1000 kelahiran hidup. Kementerian Kesehatan Republik Indonesia telah menyusun Indikator Program Kesehatan Masyarakat dalam Rencana Pembangunan Jangka Menengah Nasional (RPJMN) dan Rencana Strategi (Renstra) untuk tahun 2020-2024. Terdapat 26 indikator Program Kesehatan Masyarakat pada RPJMN tahun 2020-2024 dengan salah satu indikatornya adalah AKB serta indikator lain yang terkait AKB antara lain jumlah kematian bayi, jumlah bayi yang mendapatkan imunisasi dasar lengkap, jumlah ibu hamil yang Kekurangan Energi Kronik (KEK) dan mendapat kunjungan, jumlah dokter status keaktifan mayoritas warga desa yang berpartisipasi dalam kegiatan posyandu, ketersediaan rumah sakit bersalin terdekat dari desa, ketersediaan praktek bidan, dan jumlah kematian ibu melahirkan.

Provinsi Jawa Barat adalah provinsi dengan jumlah penduduk terbanyak di Indonesia tahun 2020 yaitu sebesar 48.274.162 dengan perbandingan antara laki-laki dan perempuan yang hampir seimbang. Jika jumlah kasus kematian bayi di Jawa Barat paling tinggi di Indonesia tidak langsung menunjukkan bahwa AKB di Jawa Barat juga tertinggi.

Berdasarkan latar belakang yang telah diuraikan, maka judul yang dikaji dalam penelitian ini adalah “Komparasi Regresi Zero Inflated Poisson (ZIP) Dan Regresi Zero Inflated Negative Binomial (ZINB) Pada Data Kematian Bayi di Jawa Barat”. Tujuan pada

penelitian ini adalah untuk menentukan model regresi ZIP dan model regresi ZINB serta untuk mengetahui model regresi yang lebih baik dalam kasus kematian bayi Provinsi Jawa Barat.

**B. Metodologi Penelitian**

Tahapan yang dilakukan pada penelitian ini antara lain dijelaskan sebagai berikut: Membentuk model Regresi Poisson, Melakukan identifikasi multikolinearitas, Melakukan pengujian overdispersi pada Regresi Poisson yang terbentuk, Membentuk model Regresi Zero Inflated Poisson (ZIP), Membentuk model Regresi Zero Inflated Negative Binomial (ZINB), Menentukan model terbaik dengan membandingkan nilai AIC.

**C. Hasil dan Pembahasan**

**Data Penelitian**

Data yang digunakan pada penelitian ini merupakan data sekunder mengenai jumlah kematian bayi pada tingkat desa/kelurahan di Jawa Barat tahun 2021. Data tersebut bersumber dari laman <https://opendata.jabarprov.go.id>. Jumlah desa/kelurahan yang menjadi objek penelitian adalah 5312 desa/kelurahan. Terdapat 4300 desa/kelurahan (81%) yang tidak memiliki jumlah kejadian kematian bayi, sedangkan sejumlah 1012 desa/kelurahan (19%) memiliki jumlah kejadian kematian bayi. Variabel respon penelitian ini adalah jumlah kematian bayi pada tingkat desa/kelurahan di Jawa Barat (Y) sedangkan variabel prediktor adalah jumlah bayi yang mendapatkan imunisasi dasar lengkap (X1), jumlah ibu hamil yang kekurangan energi kronik (KEK) dan mendapat kunjungan (X2), jumlah dokter (X3), status keaktifan mayoritas warga desa yang berpartisipasi dalam kegiatan posyandu (X4), ketersediaan Rumah Sakit bersalin terdekat dari desa (X5), ketersediaan praktek bidan (X6), dan jumlah kematian ibu melahirkan (X7). Data peubah penelitian disajikan pada Tabel 1.

Tabel 1. Peubah Penelitian

Peubah	Deskripsi	Tipe Data
Y	Jumlah kematian bayi	Kuantitatif diskrit
X <sub>1</sub>	Jumlah bayi yang mendapatkan imunisasi dasar lengkap	Kuantitatif diskrit
X <sub>2</sub>	Jumlah ibu hamil yang kekurangan energi kronik (KEK) dan mendapat kunjungan	Kuantitatif diskrit
X <sub>3</sub>	Jumlah dokter	Kuantitatif diskrit
X <sub>4</sub>	Status keaktifan mayoritas warga desa yang berpartisipasi dalam kegiatan posyandu	Kategorik biner
X <sub>5</sub>	Ketersediaan Rumah Sakit bersalin terdekat dari desa	Kategorik biner
X <sub>6</sub>	Ketersediaan praktek bidan	Kategorik biner
X <sub>7</sub>	Jumlah kematian ibu melahirkan	Kuantitatif diskrit

**Analisis Statistik Deskriptif**

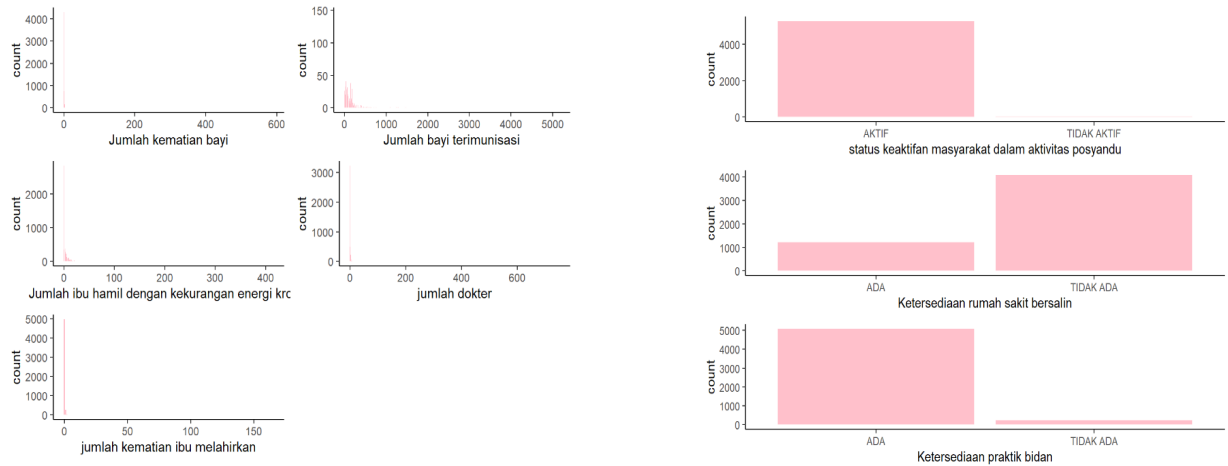
Sub bagian ini memuat statistik deskriptif peubah numerik dan histogram untuk peubah respon dan tujuh peubah prediktor. Tabel 2 menjelaskan statistik deskriptif peubah numerik.

Tabel 2. Statistik deskriptif peubah numerik

Keterangan	Peubah prediktor numerik				
	Y	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>6</sub>
Rata-rata	0.4302	126.3	4.506	1.078	0.1611
Minimum	0	0	0	0	0

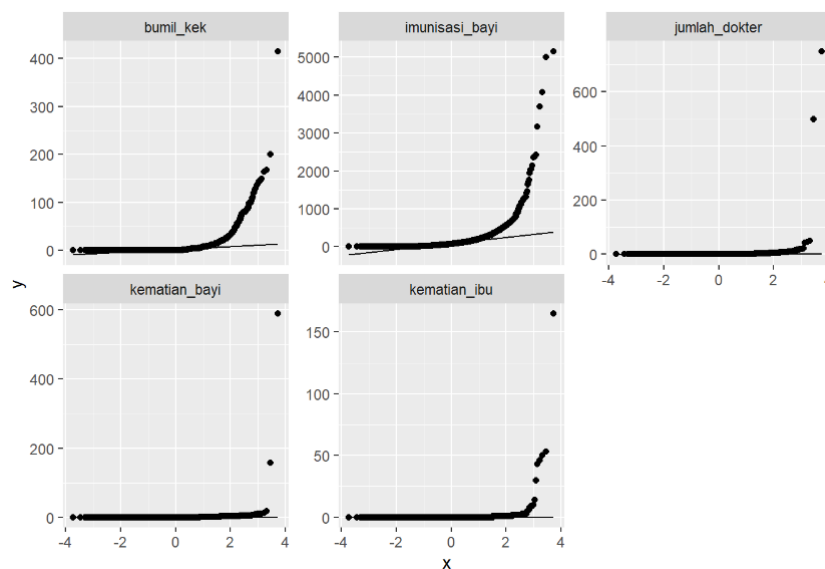
Maksimum 590 5157 415 750 165

Gambar 1 dan Gambar 2 merupakan histogram dari masing-masing empat peubah numerik dan tiga peubah kategorik.



Gambar 1. Histogram peubah numerik      Gambar 2. Histogram peubah kategorik

Berdasarkan gambar, diketahui bahwa pada tahun 2021 di Provinsi Jawa Barat status keaktifan masyarakat dalam aktivitas posyandu (X4) 5278 masyarakat dengan status aktif dan tidak aktif 34, ketersediaan rumah sakit bersalin (X5) tersedia 1225 dan tidak tersedia 4087, dan untuk ketersediaan praktik bidan(X7) 5080 ada praktik dan tidak ada praktik 232. Selanjutnya dilakukan pemeriksaan pada normalitas data menggunakan Quantile-Quantile plot yang ditunjukkan pada Gambar 3.



Gambar 3. Q-Q Plot

Berdasarkan Gambar 3 diketahui bahwa untuk pencarian data masing-masing peubah numerik pada peubah prediktor dan peubah respon tidak konsisten mengikuti garis referensinya sedemikian sehingga ada kemungkinan masing-masing peubah numerik pada peubah prediktor dan respon tidak mengikuti sebaran normal.

Pemeriksaan overdispersi dilakukan dengan program R studio dan diperoleh nilai dispersi 5004261 >1 sehingga menunjukkan adanya overdispersi. Selanjutnya dilakukan pemeriksaan faktor penyebab overdispersi yakni indikasi adanya excess zero pada peubah respon, hasil eksplorasi data pada peubah respon terdapat adanya kejadian kasus kematian bayi = 0 sebanyak 4300 amatan dari banyaknya keseluruhan amatan 5312. Data pada peubah respon didominasi amatan 0 sebesar 80.9%. Karena terdapat indikasi overdispersi dan excess zero, sehingga untuk mengatasi kasus jumlah kematian bayi digunakan pemodelan Zero Inflated Poisson dan Zero Inflated Negative Binomial (ZINB).

**Zero Inflated Poisson (ZIP)**

Model regresi ZIP bertujuan untuk memperbaiki model yang mengalami overdispersi dan memiliki banyak nilai nol pada peubah respon. Adapun pendugaan parameter dilakukan dengan menggunakan software Rstudio sehingga diperoleh hasil yang ditunjukkan pada Tabel 3:

Tabel 3. Estimasi Parameter Model Regresi ZIP

Parameter	Estimasi	SE	p-value	Signifikansi
$\beta_0$	0.7903478	0.1117564	1.53e-12	Signifikan
$\beta_1$	-0.0004731	0.0002140	0.027	Signifikan
$\beta_2$	-0.0055943	0.0038745	0.149	Tidak Signifikan
$\beta_3$	-0.0096504	0.0091031	0.289	Tidak Signifikan
$\beta_4$	-0.4458687	0.4072205	0.274	Tidak Signifikan
$\beta_5$	0.6793304	0.1127048	1.66e-09	Signifikan
$\beta_6$	-0.0272399	0.0167669	0.104	Signifikan
$\beta_7$	-0.2245376	0.1845441	0.224	Tidak Signifikan

Berdasarkan hasil pengolahan data pada Tabel 3 dapat dituliskan model ZIP dengan fungsi hubung logit

$$\text{logit}(p_i) = 0.7903478 - 0.0004731X_1 - 0.0055943X_2 - 0.0096504X_3 - 0.4458687X_4 + 0.6793304X_5 - 0.0272399X_6 - 0.2245376X_7$$

Tabel 3 menunjukkan faktor yang memberikan pengaruh terhadap kematian bayi di Provinsi Jawa Barat adalah Ibu hamil yang mengalami kekurangan energi kronik (KEK) dan mendapat kunjungan, ketersediaan Rumah sakit bersalin, dan jumlah kematian ibu.

**Zero Inflated Negative Binomial (ZINB)**

Analisis regresi ZINB digunakan sebagai pembanding dari regresi ZIP dengan tujuan untuk mengetahui model regresi mana yang baik digunakan dalam menganalisis data yang menyebabkan kematian bayi. Pendugaan parameter model regresi ZINB ditunjukkan pada Tabel 4.

Tabel 4. Estimasi Parameter Model Regresi ZINB

Parameter	Estimasi	SE	p-value	Signifikansi
$\beta_0$	-1.165e+00	5.270e-01	0.02708	Signifikan
$\beta_1$	-5.282e-04	3.955e-04	0.18166	Tidak Signifikan
$\beta_2$	-2.612e-02	1.605e-02	0.10375	Signifikan
$\beta_3$	-4.232e-02	6.960e-02	0.54317	Tidak Signifikan
$\beta_4$	2.833e-01	8.907e-01	0.7504	Tidak Signifikan
$\beta_5$	1.289e+00	4.824e-01	0.00756	Signifikan
$\beta_6$	-1.526e+01	2.734e+02	0.95549	Tidak Signifikan

Berdasarkan hasil pengolahan data pada Tabel 4 didapatkan keseluruhan model ZINB untuk model log dan logit sebagai berikut

$$\text{logit}(p_i) = -1.165 - 0.0005282X_1 - 0.02612X_2 - 0.04232X_3 + 0.2833X_4 + 1.289X_5 - 15.26X_6 - 0.4021X_7$$

### Pemilihan Model Terbaik

Pemilihan model terbaik yang dihasilkan oleh regresi Zero Inflated Poisson (ZIP) dan Zero Inflated Negative Binomial (ZINB) dapat ditentukan berdasarkan nilai Akaike's Information Criterion (AIC) dari masing-masing model. Semakin kecil nilai AIC yang dihasilkan model, model yang dihasilkan semakin baik. Tabel 5 menunjukkan nilai AIC untuk kedua model ZIP dan ZINB.

Tabel 5. Nilai AIC

Kriteria Uji	ZIP	ZINB
AIC	14373.629	7732.040

Berdasarkan Tabel 5, diketahui nilai AIC dari regresi ZINB lebih kecil dibandingkan dengan regresi ZIP. Hal tersebut menunjukkan bahwa model regresi ZINB lebih baik digunakan pada data yang menyebabkan kematian bayi berdasarkan desa/kecamatan di Provinsi Jawa Barat tahun 2021.

### Likelihood Ratio Test

Untuk menguji kelayakan model ZINB dilakukan pengujian secara serentak (overall) dengan hipotesis :  $H_0$  : Semua peubah prediktor pada model tidak mempengaruhi jumlah kematian bayi ( $\beta_1 = \dots = \beta_j$ ) dimana j adalah peubah prediktor) vs  $H_1$  : Paling sedikit ada satu peubah prediktor pada model yang mempengaruhi jumlah kematian bayi (paling sedikit ada satu  $\beta_j \neq 0$  dimana  $j = 1, 2, \dots, k$ ).

Statistik yang digunakan untuk pengujian secara serentak adalah dengan menggunakan Likelihood Ratio Test (LRT). Dengan menggunakan software R didapat nilai p-value :  $< 2.2e-16$ . Kesimpulannya adalah menolak  $H_0$ , yang berarti bahwa minimal ada satu peubah prediktor yang berpengaruh terhadap peubah respon dalam pemodelan ZINB jumlah kematian bayi desa/kecamatan di Jawa Barat tahun 2021.

#### D. Penutup

Pemodelan kasus jumlah kematian bayi berdasarkan desa/kecamatan di Jawa Barat tahun 2021 menggunakan model regresi ZIP dan ZINB dituliskan sebagai berikut.

##### 1. Model regresi ZIP

$$\logit(p_i) = 0.7903478 - 0.0004731X_1 - 0.0055943X_2 - 0.0096504 X_3 - 0.4458687X_4 + 0.6793304X_5 - 0.0272399X_6 - 0.2245376X_7$$

dengan nilai AIC 14373.629

##### 2. Model regresi ZINB

$$\logit(p_i) = -1.165 - 0.0005282X_1 - 0.02612X_2 - 0.04232X_3 + 0.2833X_4 + 1.289X_5 - 15.26X_6 - 0.4021X_7$$

dengan nilai AIC 7732.040

Berdasarkan kriteria nilai AIC dengan membandingkan model regresi ZIP dan model regresi ZINB dapat disimpulkan bahwa model regresi ZINB lebih baik digunakan untuk memodelkan jumlah kematian bayi di Jawa Barat yang mengandung banyak nilai nol dan mengalami overdispersi.

Dari model terbaik sehingga dapat diinterpretasikan bahwa Jumlah ibu hamil dengan kekurangan energi kronik akan meningkatkan kematian bayi sebesar 1.026464 setiap satu satuan peubah predictor. Dan desa yang memiliki ketersediaan rumah sakit bersalin akan meningkatkan kematian bayi sebesar 3.629156 dari pada desa yang tidak memiliki ketersediaan bersalin

#### Daftar Pustaka

- Badan Pusat Statistik. 2021. Sensus Penduduk 2020. Jakarta Pusat : Badan Pusat Statistik.
- CIA. 2020. CIA World Factbook. <https://www.indexmundi.com/map/?v=29&r=as&l=en>. Diakses pada 1 September 2025.
- Direktorat Jenderal Kesehatan Masyarakat Indonesia. Kementerian Kesehatan RI. 2020. Pedoman Indikator Program Kesehatan Masyarakat Dalam RPJMN dan Renstra Kementerian Kesehatan Tahun 2020-2024.
- Garay AM, Hashimoto EM. 2011. On Estimation And Influence Diagnostics for Zero Inflated Negative Binomial Regression Model. *Computational Statistics & Data Analysis*. 55(3):1304-1318.
- Hilbe JM. 2011. *Negative Binomial Regression*. Second Edition. New York: Cambridge University Press.
- Hosmer DW, Lemeshow S. 2000. *Applied Logistic Regression*. New York: John Wiley & Sons, Inc.
- Ismail N, Jemain AA. 2005. Generalized Poisson Regression: An Alternative For Risk Classification. *Jurnal Teknologi*. 43(1):39-50.
- Jansakul N, Hinde JP. 2002. Score Tests for Zero-Inflated Poisson Models. *Computational Statistics & Data Analysis*. 40(1):75-96.
- Lambert D. 1992. Zero Inflated Poisson Regression, with an Application to Detect in Manufacturing. *Technometrics*. 34(1):1-14.
- Myers RH, Douglas CM, Vining GG, Timothy JR. 2010. *Generalized Linear Models with Applications in Engineering and The Sciences*. Second edition. New Jersey: John Wiley and Sons.
- Open Data Jabar. 2021. <https://opendata.jabarprov.go.id>. Diakses pada 1 September 2025.

- Purnama DI, 2021. Comparison of Zero Inflated Poisson (ZIP) Regression, Zero Inflated Negative Binomial Regression (ZINB) and Binomial Negative Hurdle Regression (HNB) to Model Daily Cigarette Consumption Data for Adult Population in Indonesia. *J Matematika, Stat, Komputasi*. 17(3):357-269.
- Ruliana, Hendikawati P, Agoestanto A. 2016. Pemodelan *Generalized Poisson Regression* (GPR) Untuk Mengatasi Pelanggaran *Equidispersi* Pada Regresi *Poisson* Kasus Campak di Kota Semarang Tahun 2013. *UNNES Journal of Mathematics*, 5(1) : 39-46.
- Samson, Rizki SW, Perdana H. 2022. Perbandingan Regresi Zero Inflated Poisson (ZIP) dan Regresi Zero Inflated Negative Binomial (ZINB) Pada Data Overdispersi. *Bimaster*. 11(1):103-110