

Ketidakesesuaian Distribusi *Poisson* dalam Pemodelan Jumlah Kehadiran Siswa Sekolah Menengah Kejuruan

Inadequacy of the Poisson Distribution in Modeling Vocational High School Student Attendance Counts

Aditya Pranata^{1*}, Mitra Pranata, Marsini Sibuea, Sumiran, Joni, Rian Farta Wijaya

¹Fakultas Pascasarjana, Universitas Pembangunan Panca Budi Medan, Medan, Indonesia

Article Info

Article history:

Received Mei 22, 2026

Revised Agustus 28, 2026

Accepted Agustus 28, 2026

Published Agustus 31, 2026

Kata Kunci:

Distribusi Poisson;
Underdispersion;
Data Cacahan Berbatas;
Uji Kesesuaian;
Kehadiran Siswa

Keywords:

Poisson Distribution;
Underdispersion;
Bounded Count Data;
Goodness-of-Fit;
Student Attendance

License: [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)



Copyright: © 2026 The Author(s)

ABSTRAK

Kehadiran siswa merupakan indikator penting dalam kontinuitas proses pembelajaran, tetapi jumlah siswa hadir per hari pada kelas berukuran tetap merupakan data cacahan dengan batas atas yang dapat membatasi kesesuaian model Poisson. Penelitian ini bertujuan mengevaluasi kelayakan distribusi Poisson dalam memodelkan jumlah kehadiran harian 30 siswa kelas XI TKJ SMK Putra Anda Binjai selama 224 hari efektif pada periode Juli 2025-April 2026. Parameter rata-rata diestimasi dengan *Maximum Likelihood Estimation (MLE)*, sedangkan kesesuaian model dievaluasi melalui rasio variansi terhadap mean, probabilitas model pada nilai yang melampaui batas kelas, serta statistik *Kolmogorov-Smirnov (K-S)* yang dikalibrasi dengan bootstrap parametrik 20.000 replikasi. Hasil menunjukkan $\lambda = 28,5134$ siswa/hari atau tingkat kehadiran agregat 95,04%, dengan variansi 0,4220 dan rasio variansi/mean 0,0148 yang menunjukkan underdispersion sangat kuat. Model Poisson juga menempatkan sekitar 34,50% probabilitas pada $X > 30$, yaitu kejadian yang tidak mungkin untuk kelas berukuran 30 siswa. Statistik K-S sebesar $D = 0,4367$ dengan $p\text{-value bootstrap} < 0,0001$ menunjukkan bahwa data tidak konsisten dengan distribusi Poisson. Dengan demikian, distribusi Poisson tidak tepat digunakan secara langsung untuk pemodelan probabilitas kehadiran pada data cacahan berbatas ini; model dengan batas atas eksplisit dan/atau kemampuan mengakomodasi underdispersion perlu dibandingkan pada penelitian lanjutan. Temuan ini terbatas pada satu kelas dan satu sekolah.

ABSTRACT

Student attendance is important for continuity of learning, but the daily number of students present in a fixed-size class is bounded count data, which can conflict with the structural properties of a Poisson model. This study evaluates the suitability of the Poisson distribution for modeling daily attendance counts of 30 Class XI TKJ students at SMK Putra Anda Binjai over 224 effective school days from July 2025 to April 2026. The mean parameter was estimated by Maximum Likelihood Estimation (MLE), while model fit was assessed using the variance-to-mean ratio, the probability assigned beyond the class-size bound, and a Kolmogorov-Smirnov (K-S) statistic calibrated by a 20,000-replication parametric bootstrap. The estimated rate was $\lambda = 28.5134$ students/day, corresponding to an aggregate attendance rate of 95.04%. The empirical variance was 0.4220, producing a variance-to-mean ratio of 0.0148 and indicating very strong underdispersion. The fitted Poisson model also assigned approximately 34.50% probability to $X > 30$, an impossible

event for a class of 30 students. The K-S statistic was $D = 0.4367$ with a bootstrap p-value < 0.0001 , indicating that the data are inconsistent with the Poisson distribution. Therefore, direct Poisson modeling is not appropriate for these bounded attendance counts, and models with explicit upper bounds and/or flexible dispersion should be compared in future work. These findings are limited to one class in one school.

Corresponding Author:

Aditya Pranata
Fakultas Pascasarjana, Universitas Pembangunan Panca Budi Medan,
Medan, Indonesia
Email: adityapranata1@gmail.com

1. PENDAHULUAN

Kehadiran siswa merupakan indikator penting dalam kontinuitas proses pembelajaran dan pencapaian akademik. Ketidakhadiran yang berulang dapat mengurangi paparan siswa terhadap pembelajaran, dan penelitian berbasis data pendidikan menunjukkan bahwa absensi berkaitan dengan capaian akademik yang lebih rendah serta risiko hasil pendidikan jangka panjang [1]-[4]. Pada konteks pembelajaran, kehadiran kelas juga berasosiasi dengan performa akademik, meskipun hubungan tersebut tidak serta-merta menunjukkan hubungan kausal [5]. Oleh karena itu, analisis pola kehadiran berbasis data diperlukan untuk menyediakan gambaran kuantitatif yang dapat mendukung pengelolaan pembelajaran dan pemantauan kehadiran siswa.

Literatur mutakhir mengenai kehadiran siswa lebih banyak membahas konsekuensi ketidakhadiran, faktor risiko, dan hubungan kehadiran dengan prestasi [1]-[5]. Kajian-kajian tersebut memberikan dasar kuat mengenai pentingnya kehadiran, tetapi umumnya memperlakukan kehadiran sebagai prediktor atau luaran pendidikan, bukan sebagai variabel cacahan yang bentuk distribusi probabilitasnya perlu dievaluasi. Dalam kelas berukuran tetap, jumlah siswa hadir per hari hanya dapat mengambil nilai dari 0 sampai jumlah siswa terdaftar. Karakteristik batas atas tersebut penting karena model cacahan yang tidak memiliki batas atas dapat memberikan probabilitas pada nilai yang secara substantif tidak mungkin terjadi [6].

Distribusi Poisson merupakan model acuan untuk data cacahan dengan karakteristik equidispersion, yaitu nilai harapan dan variansi yang sama. Namun, data cacahan empiris dapat menunjukkan overdispersion maupun underdispersion, sehingga asumsi Poisson perlu diuji dan tidak dapat diterima hanya karena variabel berbentuk cacahan [7]-[10]. Pada data yang memiliki batas atas alami, pemilihan model juga harus mempertimbangkan dukungan distribusi (support), bukan hanya kecocokan rata-rata [6]. Dengan demikian, distribusi Poisson dalam penelitian ini diposisikan sebagai model kandidat yang dievaluasi secara kritis, bukan sebagai model yang diasumsikan sesuai sejak awal.

Kesenjangan penelitian terletak pada masih terbatasnya evaluasi kesesuaian distribusi Poisson untuk jumlah kehadiran harian pada kelas dengan ukuran tetap, terutama ketika dua sumber ketidaksesuaian dapat muncul secara bersamaan, yaitu underdispersion dan batas atas jumlah siswa. Literatur kehadiran yang digunakan dalam penelitian ini terutama membahas konsekuensi absensi dan hubungan kehadiran dengan prestasi [1]-[5], sedangkan literatur metodologis membahas data cacahan terbatas [6], underdispersion [7]-[10], serta pengujian goodness-of-fit untuk distribusi diskrit [11]-[15]. Kombinasi ketiga persoalan tersebut pada data kehadiran kelas berukuran tetap belum menjadi fokus utama dalam rujukan yang digunakan. Selain itu, pengujian Kolmogorov-Smirnov untuk distribusi diskrit dengan parameter yang diestimasi dari data memerlukan kalibrasi yang sesuai dan tidak seharusnya langsung menggunakan acuan uji K-S kontinu standar [11]-[15]. Kebaruan penelitian ini adalah mengevaluasi model Poisson melalui bukti yang saling melengkapi, yaitu rasio variansi terhadap mean, besarnya probabilitas model yang berada di luar batas kelas, statistik K-S yang dikalibrasi dengan bootstrap parametrik, serta pembandingan terbatas berdasarkan karakteristik model Binomial.

Berdasarkan kesenjangan tersebut, penelitian ini bertujuan mengestimasi parameter rata-rata kehadiran, menghitung probabilitas Poisson pada nilai pengamatan, menguji kesesuaian model, dan merumuskan implikasi pemodelan yang lebih tepat bagi data kehadiran dengan ukuran kelas tetap.

2. METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan analisis probabilitistik dan pengujian kesesuaian distribusi (goodness-of-fit). Fokus analisis bukan sekadar mendeskripsikan tingkat kehadiran, tetapi mengevaluasi apakah distribusi Poisson layak digunakan untuk memodelkan jumlah siswa yang hadir pada setiap hari efektif pembelajaran.

Penelitian dilaksanakan di SMK Putra Anda Binjai, Jalan W. R. Mongonsidi No. 22, Kota Binjai, Provinsi Sumatera Utara. Kohort yang diamati terdiri atas seluruh 30 siswa kelas XI TKJ (Teknik Komputer dan Jaringan) pada tahun ajaran 2025/2026. Unit analisis adalah jumlah siswa yang hadir pada setiap hari efektif pembelajaran. Periode pengamatan mencakup 10 bulan, yaitu Juli 2025 hingga April 2026, dengan total 224 hari efektif; karena seluruh hari efektif pada periode tersebut dianalisis, tidak dilakukan pengambilan sampel hari pengamatan. Analisis dan pelaporan dilakukan pada tingkat agregat harian, sehingga nama maupun identitas individual siswa tidak menjadi variabel analisis dan tidak ditampilkan dalam hasil penelitian.

Data yang digunakan merupakan data sekunder berupa rekapitulasi presensi sekolah yang diperoleh melalui dokumentasi. Status kehadiran harian setiap siswa terdiri atas hadir (H), tidak hadir tanpa keterangan/alpha (A), izin (I), dan sakit (S). Data akhir yang dianalisis mencakup 224 hari efektif sebagaimana tercatat pada rekap administrasi sekolah. Variabel analisis adalah jumlah siswa hadir per hari, X , yang dihitung dengan $X = N - (nA + nI + nS)$, dengan $N = 30$ sebagai ukuran kelas, sedangkan nA , nI , dan nS masing-masing adalah jumlah siswa alpha, izin, dan sakit pada hari yang sama. Konsistensi data sekunder diperiksa melalui pencocokan rekap administrasi dengan lembar absensi harian guru dan rekap bulanan Bimbingan dan Konseling (BK) sekolah. Hari di luar daftar hari efektif tidak dimasukkan ke dalam unit analisis.

Analisis dimulai dengan mengubah data presensi per siswa menjadi jumlah siswa hadir untuk setiap hari efektif. Parameter rata-rata distribusi Poisson kemudian diestimasi menggunakan Maximum Likelihood Estimation (MLE) sebagai rata-rata jumlah kehadiran harian:

$$\hat{\lambda} = (\sum x_i) / n$$

di mana $\hat{\lambda}$ adalah estimasi rata-rata jumlah siswa hadir per hari, x_i adalah jumlah siswa hadir pada hari ke- i , dan $n = 224$ adalah jumlah hari efektif. Kesesuaian terhadap asumsi equidispersion diperiksa dengan membandingkan mean ($\bar{x}$) dan variansi (s^2) melalui rasio variansi terhadap mean, $VMR = s^2/\bar{x}$. Nilai VMR yang mendekati 1 konsisten dengan equidispersion, sedangkan $VMR < 1$ menunjukkan underdispersion [7]-[10]. Selain dispersi, dilakukan pemeriksaan struktural terhadap dukungan distribusi. Data aktual dibatasi pada $X \leq 30$, sementara distribusi Poisson memiliki dukungan tanpa batas atas; karena itu, probabilitas $P(X > 30)$ dihitung untuk mengukur massa probabilitas model pada kejadian yang mustahil secara substantif [6]. Sebagai pembanding terbatas, digunakan karakteristik Binomial dengan $n = 30$ dan $\hat{p} = \bar{x}/30$. Model ini memiliki mean $n\hat{p}$ dan variansi $n\hat{p}(1-\hat{p})$. Perbandingan variansi empiris dengan variansi Binomial digunakan sebagai diagnostik tambahan untuk menilai apakah pemberian batas atas saja sudah cukup menjelaskan tingkat dispersi data.

Probabilitas teoretis untuk setiap jumlah kehadiran dihitung menggunakan fungsi massa peluang distribusi Poisson:

$$P(X = x) = \exp(-\lambda) \lambda^x / x!, \text{ untuk } x = 0, 1, 2, \dots$$

dengan $e \approx 2,71828$, λ sebagai parameter rata-rata yang diestimasi, dan $x!$ sebagai faktorial dari nilai pengamatan. Kesesuaian distribusi selanjutnya dievaluasi menggunakan statistik Kolmogorov-Smirnov (K-S) yang membandingkan fungsi distribusi kumulatif empiris dengan fungsi distribusi kumulatif Poisson terpasang:

$$D = \max |S_n(x) - F_0(x; \lambda)|$$

di mana $S_n(x)$ adalah fungsi distribusi kumulatif empiris dan $F_0(x; \lambda)$ adalah fungsi distribusi kumulatif Poisson dengan parameter yang diestimasi dari data. Karena data bersifat diskrit dan λ diperoleh dari sampel yang sama, p-value K-S kontinu standar tidak digunakan sebagai dasar keputusan [12], [13]. Kalibrasi dilakukan melalui bootstrap parametrik [15] dengan 20.000 replikasi: pada setiap replikasi dibangkitkan sampel Poisson berukuran 224 dari λ , parameter diestimasi kembali, statistik D dihitung ulang, kemudian p-value bootstrap ditentukan dari proporsi statistik bootstrap yang sekurang-kurangnya sebesar D observasi. Hipotesis nol menyatakan bahwa distribusi jumlah kehadiran konsisten dengan model Poisson, dan H_0 ditolak jika p-value bootstrap $< \alpha = 0,05$. Prosedur ini melengkapi evaluasi dispersi dan batas dukungan sehingga keputusan kesesuaian tidak hanya bergantung pada satu indikator. Pemeriksaan autokorelasi antarhari belum dimasukkan dalam analisis; karena itu, hasil bootstrap diinterpretasikan sebagai evaluasi kesesuaian distribusi marginal, sedangkan kemungkinan ketergantungan temporal antarhari diperlakukan sebagai salah satu keterbatasan penelitian. Analisis dapat direplikasi menggunakan lingkungan komputasi statistik R [16].

Secara keseluruhan, alur analisis dimulai dari verifikasi dan rekapitulasi data presensi harian, dilanjutkan dengan estimasi λ , pemeriksaan VMR dan batas dukungan, perhitungan probabilitas Poisson, perbandingan karakteristik variansi dengan model Binomial terbatas, serta pengujian K-S yang dikalibrasi dengan bootstrap parametrik. Hasil diagnostik tersebut kemudian diinterpretasikan secara terpadu untuk menilai kelayakan penggunaan distribusi Poisson pada data kehadiran kelas berukuran tetap.

3. HASIL DAN PEMBAHASAN

Data penelitian terdiri atas 224 unit pengamatan harian dari satu kelas berukuran tetap 30 siswa. Dengan demikian, terdapat 6.720 siswa-hari yang secara teoritis dapat hadir selama periode pengamatan. Rekap administrasi menunjukkan 6.387 siswa-hari hadir dan 333 siswa-hari tidak hadir. Ringkasan bulanan digunakan untuk menggambarkan pola deskriptif, sedangkan mean, variansi, VMR, probabilitas model, dan pengujian distribusi dihitung dari unit pengamatan harian. Pemisahan ini penting agar rekap bulanan tidak diperlakukan sebagai dasar langsung untuk pengujian kesesuaian distribusi.

Rekapitulasi kehadiran siswa kelas XI TKJ per bulan selama periode Juli 2025-April 2026 disajikan pada Tabel 1.

Tabel 1. Rekapitulasi Data Kehadiran Siswa Kelas XI TKJ Per Bulan Tahun Ajaran 2025/2026

No	Bulan	Semester	Hari Efektif	Ketidakhadiran (siswa-hari)	Kehadiran (siswa-hari)	Rata-rata Hadir/Hari
1	Juli 2025	Ganjil	23	22	668	29,04
2	Agustus 2025	Ganjil	23	23	667	29,00
3	September 2025	Ganjil	22	23	637	28,95
4	Oktober 2025	Ganjil	23	46	644	28,00
5	November 2025	Ganjil	22	57	603	27,41
6	Desember 2025	Ganjil	23	33	657	28,57
7	Januari 2026	Genap	23	57	633	27,52
8	Februari 2026	Genap	20	30	570	28,50
9	Maret 2026	Genap	23	19	671	29,17
10	April 2026	Genap	22	23	637	28,95

No	Bulan	Semester	Hari Efektif	Ketidakhadiran (siswa-hari)	Kehadiran (siswa-hari)	Rata-rata Hadir/Hari
Total			224	333	6.387	28,51

Sumber: Arsip Administrasi Presensi SMK Putra Anda Binjai, 2025/2026

Berdasarkan Tabel 1, ketidakhadiran tertinggi tercatat pada November 2025 dan Januari 2026, masing-masing 57 siswa-hari, sedangkan ketidakhadiran terendah terjadi pada Maret 2026 sebanyak 19 siswa-hari. Rata-rata kehadiran bulanan berkisar antara 27,41 siswa/hari pada November 2025 dan 29,17 siswa/hari pada Maret 2026. Pola tersebut menunjukkan adanya variasi kehadiran antarbulan, tetapi desain penelitian ini tidak mengukur penyebab variasi tersebut. Oleh karena itu, perbedaan antarbulan tidak diatribusikan pada periode ujian, libur semester, motivasi, atau faktor lain tanpa data penjelas yang secara langsung diukur.

Parameter rata-rata kehadiran berdasarkan seluruh 224 hari efektif adalah $\bar{x} = 6.387/224 = 28,5134 \approx 28,51$ siswa/hari. Mean empiris sebesar 28,5134 dan variansi sebesar 0,4220 menghasilkan $VMR = 0,4220/28,5134 = 0,0148$. Nilai yang jauh di bawah 1 menunjukkan underdispersion yang sangat kuat dan bertentangan dengan asumsi equidispersion Poisson [7]-[10]. Secara deskriptif, tingkat kehadiran agregat mencapai $6.387/6.720 \times 100\% = 95,04\%$. Tingginya persentase ini menunjukkan bahwa jumlah kehadiran terkonsentrasi dekat dengan batas maksimum 30 siswa, yang sekaligus membatasi variasi data harian.

Dengan $\bar{x} = 28,5134$, probabilitas teoretis Poisson untuk $x = 24$ sampai $x = 30$ dihitung dan disajikan pada Tabel 2. Rentang tersebut dipilih untuk menampilkan bagian distribusi di sekitar rata-rata kehadiran sekaligus mendekati batas atas kelas, karena area inilah yang paling relevan untuk melihat konsentrasi probabilitas dan ketidaksesuaian dukungan model.

Tabel 2. Probabilitas Distribusi Poisson Kehadiran Siswa Kelas XI TKJ ($x = 24$ s.d. 30)

Jumlah Siswa Hadir (x)	Probabilitas $P(X = x)$	$P(X = x)$ dalam (%)	Keterangan
24	0,055624	5,5624%	
25	0,063441	6,3441%	
26	0,069573	6,9573%	
27	0,073473	7,3473%	
28	0,074820	7,4820%	Nilai tertinggi (modus)
29	0,073565	7,3565%	
30	0,069919	6,9919%	

Sumber: Hasil pengolahan data penelitian, 2026

Berdasarkan Tabel 2, probabilitas Poisson terbesar pada rentang yang ditampilkan terjadi pada $x = 28$ dengan $P(X = 28) = 0,074820$ atau 7,4820%, konsisten dengan modus distribusi Poisson untuk $\bar{x} = 28,5134$. Probabilitas $P(X = 30)$ adalah 0,069919 atau 6,9919%, sedangkan $P(28 \leq X \leq 30) = 21,8304\%$. Namun, probabilitas kumulatif $P(X \geq 28)$ menurut model mencapai sekitar 56,33%; selisihnya menunjukkan bahwa sekitar 34,50% massa probabilitas Poisson berada pada $X > 30$. Nilai tersebut merepresentasikan kejadian yang mustahil karena ukuran kelas hanya 30 siswa. Dengan demikian, selain underdispersion, terdapat ketidaksesuaian struktural antara dukungan distribusi Poisson yang tidak terbatas atas dan dukungan data kehadiran yang dibatasi oleh ukuran kelas [6].

Sebagai pembanding terbatas, $\hat{p}$ dihitung dari $28,5134/30 = 0,9504$. Dengan $n = 30$, model Binomial memiliki mean teoretis yang sama, yaitu sekitar 28,5134 siswa, tetapi variansi teoretisnya

$n\hat{p}(1-\hat{p}) \approx 1,4129$. Variansi empiris 0,4220 hanya sekitar 29,87% dari variansi Binomial tersebut. Artinya, pemberian batas atas eksplisit memang menghilangkan probabilitas pada $X > 30$, tetapi belum sepenuhnya menjelaskan tingkat konsentrasi data yang sangat tinggi. Temuan ini menguatkan kebutuhan terhadap model terbatas yang sekaligus mampu menangani underdispersion, bukan hanya mengganti Poisson dengan model yang memiliki support terbatas [6]-[10]. Pengujian formal kesesuaian dilakukan menggunakan statistik K-S dengan kalibrasi bootstrap parametrik karena data bersifat diskrit dan parameter Poisson diestimasi dari data. Ringkasan hasil pengujian disajikan pada Tabel 3.

Tabel 3. Hasil Uji Kecocokan Distribusi Poisson (Kolmogorov-Smirnov)

Keterangan	Nilai
Statistik K-S (D)	0,4367
Replikasi bootstrap	20.000
P-value bootstrap	< 0,0001
Tingkat Signifikansi (α)	0,05
Keputusan	H_0 ditolak — data tidak konsisten dengan distribusi Poisson

Sumber: Hasil pengolahan dan bootstrap parametrik, 2026

Berdasarkan Tabel 3, statistik observasi adalah $D = 0,4367$. Pada 20.000 replikasi bootstrap parametrik, p-value terkalibrasi diperoleh < 0,0001, sehingga H_0 ditolak pada $\alpha = 0,05$. Hasil ini menunjukkan bahwa distribusi empiris jumlah kehadiran harian tidak konsisten dengan model Poisson terpasang. Keputusan tersebut searah dengan dua diagnostik lainnya, yaitu VMR sebesar 0,0148 yang menunjukkan underdispersion sangat kuat dan massa probabilitas sekitar 34,50% pada $X > 30$ yang berada di luar batas kelas. Karena ketiga indikator memberikan arah yang sama, ketidaksesuaian model tidak disimpulkan hanya dari uji K-S konvensional.

Temuan ini konsisten dengan literatur tentang data cacahan terbatas yang menekankan bahwa variabel dengan batas atas alami tidak selalu tepat dimodelkan menggunakan distribusi cacahan tanpa batas atas [6]. Literatur underdispersion juga menunjukkan bahwa kendala equidispersion Poisson dapat menghasilkan kecocokan yang buruk ketika variansi data jauh lebih kecil daripada mean, sehingga model yang lebih fleksibel diperlukan [7]-[10]. Hasil perbandingan Binomial pada penelitian ini memperjelas dua persoalan yang berbeda: batas dukungan dan tingkat dispersi. Binomial memperbaiki persoalan dukungan karena tidak memberi probabilitas pada $X > 30$, tetapi variansinya masih lebih besar daripada variansi empiris. Dari sisi pengujian, hasil penelitian mendukung penggunaan prosedur goodness-of-fit yang mempertimbangkan sifat diskrit distribusi dan estimasi parameter, alih-alih menerapkan acuan K-S kontinu secara langsung [11]-[15]. Dengan demikian, kontribusi utama penelitian bukan sekadar menerapkan distribusi Poisson pada data pendidikan, melainkan mendiagnosis secara empiris mengapa model tersebut tidak sesuai dan menunjukkan bahwa pemodelan data kehadiran perlu memperhatikan batas kelas sekaligus struktur dispersi.

Nilai rata-rata 28,51 siswa/hari dan tingkat kehadiran agregat 95,04% menunjukkan tingkat kehadiran yang tinggi selama periode pengamatan, tetapi temuan tersebut tidak cukup untuk menyimpulkan adanya 'budaya kehadiran' atau dampak terhadap hasil belajar. Penelitian terdahulu menunjukkan adanya hubungan antara kehadiran dan capaian akademik [1]-[5], sedangkan penelitian ini tidak mengukur nilai akademik, motivasi, kondisi keluarga, maupun faktor penyebab ketidakhadiran. Secara praktis, pemilihan model probabilitas juga penting ketika data kehadiran digunakan sebagai masukan dashboard, sistem monitoring, atau mekanisme early warning sekolah. Model Poisson yang

memberi probabilitas pada jumlah hadir lebih dari 30 dapat menghasilkan keluaran yang secara substantif tidak valid, sehingga model terbatas lebih aman digunakan sebagai dasar analitik. Namun, penelitian ini belum menguji implementasi sistem informasi atau dampak keputusan yang dihasilkan oleh model tersebut. Bulan dengan ketidakhadiran relatif lebih tinggi dapat digunakan sebagai titik awal pemantauan sekolah, tetapi penyebabnya tetap perlu diteliti dengan data individual atau desain eksplanatori tersendiri.

Implikasi metodologis penelitian adalah bahwa $\lambda = 28,51$ tetap berguna sebagai statistik deskriptif rata-rata, tetapi probabilitas prediktif dari model Poisson tidak layak digunakan secara langsung karena model gagal memenuhi karakteristik dispersi dan batas dukungan data. Penelitian lanjutan sebaiknya membandingkan model dengan batas atas eksplisit, seperti model Binomial, dengan model yang dapat mengakomodasi dispersi secara lebih fleksibel. Pembandingan Binomial pada penelitian ini masih bersifat diagnostik melalui support, mean, dan variansi, belum berupa uji kecocokan formal antar-model. Jika data harian rinci tersedia untuk analisis lanjutan, distribusi frekuensi empiris pada setiap nilai x sebaiknya dilaporkan dan perbandingan model dapat diperluas menggunakan ukuran kecocokan, kovariat individual, efek waktu, serta pola antarkelas. Selain itu, ketergantungan temporal antarkelas perlu diperiksa agar struktur probabilitas yang dipilih tidak hanya tepat secara marginal, tetapi juga sesuai dengan karakteristik runtun waktu data.

Keterbatasan penelitian ini adalah cakupan data yang hanya berasal dari satu kelas, satu sekolah, dan satu tahun ajaran. Karena itu, hasil tidak dapat langsung digeneralisasikan ke sekolah, jenjang, atau ukuran kelas lain. Analisis juga belum memasukkan pengujian ketergantungan temporal antarkelas dan belum membandingkan model alternatif secara formal menggunakan data frekuensi harian. Keterbatasan tersebut menjadi arah utama untuk pengujian lanjutan.

4. KESIMPULAN

Rata-rata kehadiran harian siswa selama 224 hari efektif adalah 28,51 dari 30 siswa, setara dengan tingkat kehadiran agregat 95,04%. Namun, variansi sebesar 0,4220 dan VMR sebesar 0,0148 menunjukkan underdispersion yang sangat kuat, sedangkan model Poisson menempatkan sekitar 34,50% probabilitas pada $X > 30$ yang tidak mungkin terjadi untuk kelas berukuran tetap. Uji K-S yang dikalibrasi melalui 20.000 replikasi bootstrap parametrik menghasilkan $D = 0,4367$ dengan $p\text{-value} < 0,0001$, sehingga data tidak konsisten dengan distribusi Poisson. Pembandingan Binomial dengan $n = 30$ dan $p = 0,9504$ menghasilkan variansi teoretis sekitar 1,4129, yang masih lebih tinggi daripada variansi empiris; hal ini menunjukkan bahwa penggunaan model terbatas saja belum sepenuhnya mengatasi underdispersion. Oleh karena itu, distribusi Poisson tidak tepat digunakan secara langsung sebagai model probabilitas kehadiran pada data ini. Nilai rata-rata tetap dapat digunakan secara deskriptif, sedangkan penelitian lanjutan perlu membandingkan secara formal model yang memiliki batas atas eksplisit dan mampu mengakomodasi underdispersion. Kesimpulan ini terbatas pada satu kelas dan satu sekolah sehingga generalisasi ke konteks lain perlu dilakukan secara hati-hati.

PERNYATAAN PENGGUNAAN AI

Penulis menyatakan bahwa dalam penelitian mengenai pemodelan distribusi Poisson ini, tidak ada bagian analisis data, penarikan kesimpulan, maupun penyusunan naskah utama yang dihasilkan menggunakan bantuan AI (Artificial Intelligence). AI generatif hanya digunakan sebatas sebagai alat bantu pengecekan tata bahasa pada penulisan draf naskah.

KONTRIBUSI PENULIS

Konseptualisasi: Aditya Pranata, Mitra Pranata
Metodologi: Aditya Pranata

Pengembangan perangkat lunak/sistem: -
Pengumpulan data: Aditya Pranata
Analisis data: Mitra Pranata, Sumiran, Joni
Validasi: Joni, Marsini
Visualisasi: -
Penelusuran literatur: Aditya Pranata, Marsini
Penulisan draf awal: Aditya Pranata, Sumiran
Penelaahan dan penyuntingan: Mitra Pranata, Marsini
Supervisi: Joni, Marsini
Administrasi penelitian: Aditya Pranata

PERNYATAAN PENDANAAN

Penelitian ini dilakukan secara mandiri oleh penulis dan tidak mendapatkan dukungan pendanaan dari institusi kampus, pihak eksternal, maupun program hibah penelitian mana pun.

KONFLIK KEPENTINGAN

Para penulis menyatakan tidak terdapat konflik kepentingan dalam penelitian dan publikasi artikel ini.

DAFTAR PUSTAKA

- [1] M. Klein, E. M. Sosu, and S. Dare, "School Absenteeism and Academic Achievement: Does the Reason for Absence Matter?," *AERA Open*, vol. 8, no. 1, pp. 1–14, 2022, doi: 10.1177/23328584211071115.
- [2] R. A. London, M. Sanchez, and S. Castrechini, "The Dynamics of Chronic Absence and Student Achievement," *Education Policy Analysis Archives*, vol. 24, art. no. 112, 2016, doi: 10.14507/epaa.24.2471.
- [3] N. E. Smerillo, A. J. Reynolds, J. A. Temple, and S.-R. Ou, "Chronic absence, eighth-grade achievement, and high school attainment in the Chicago Longitudinal Study," *Journal of School Psychology*, vol. 67, pp. 163–178, 2018, doi: 10.1016/j.jsp.2017.11.001.
- [4] J. Gubbels, C. E. van der Put, and M. Assink, "Risk Factors for School Absenteeism and Dropout: A Meta-Analytic Review," *Journal of Youth and Adolescence*, vol. 48, no. 9, pp. 1637–1667, 2019, doi: 10.1007/s10964-019-01072-5.
- [5] V. Kassarnig, A. Bjerre-Nielsen, E. Mones, S. Lehmann, and D. D. Lassen, "Class attendance, peer similarity, and academic performance in a large field study," *PLOS ONE*, vol. 12, no. 11, art. no. e0187078, 2017, doi: 10.1371/journal.pone.0187078.
- [6] C. L. Britt, M. Rocque, and G. M. Zimmerman, "The Analysis of Bounded Count Data in Criminology," *Journal of Quantitative Criminology*, vol. 34, no. 2, pp. 591–607, 2018, doi: 10.1007/s10940-017-9346-9.
- [7] K. F. Sellers and D. S. Morris, "Underdispersion models: Models that are 'under the radar'," *Communications in Statistics—Theory and Methods*, vol. 46, no. 24, pp. 12075–12086, 2017, doi: 10.1080/03610926.2017.1291976.
- [8] A. Huang, "Mean-parametrized Conway–Maxwell–Poisson regression models for dispersed counts," *Statistical Modelling*, vol. 17, no. 6, pp. 359–380, 2017, doi: 10.1177/1471082X17697749.
- [9] M. Melo and A. P. Alencar, "Conway–Maxwell–Poisson Autoregressive Moving Average Model for Equidispersed, Underdispersed, and Overdispersed Count Data," *Journal of Time Series Analysis*, vol. 41, no. 6, pp. 830–857, 2020, doi: 10.1111/jtsa.12550.
- [10] X.-J. Li, G.-L. Tian, M. Zhang, G. T. S. Ho, and S. Li, "Modeling Under-Dispersed Count Data by the Generalized Poisson Distribution via Two New MM Algorithms," *Mathematics*, vol. 11, no. 6, art. no. 1478, 2023, doi: 10.3390/math11061478.

- [11] P. A. Mijburgh and I. J. H. Visagie, “An overview of goodness-of-fit tests for the Poisson distribution,” *South African Statistical Journal*, vol. 54, no. 2, pp. 207–230, 2020, doi: 10.37920/sasj.2020.54.2.6.
- [12] A. Zeimbekakis, E. D. Schifano, and J. Yan, “On Misuses of the Kolmogorov–Smirnov Test for One-Sample Goodness-of-Fit,” *The American Statistician*, vol. 78, no. 4, pp. 481–487, 2024, doi: 10.1080/00031305.2024.2356095.
- [13] D. S. Dimitrova, V. K. Kaishev, and S. Tan, “Computing the Kolmogorov-Smirnov Distribution When the Underlying CDF is Purely Discrete, Mixed, or Continuous,” *Journal of Statistical Software*, vol. 95, no. 10, pp. 1–42, 2020, doi: 10.18637/jss.v095.i10.
- [14] S. G. Meintanis, J. Ngatchou-Wandji, L. Santana, and M. Smuts, “Goodness-of-fit tests for discrete response models with covariates,” *Statistical Papers*, vol. 66, no. 3, art. no. 54, 2025, doi: 10.1007/s00362-024-01654-x.
- [15] G. Szűcs, “Parametric bootstrap tests for continuous and discrete distributions,” *Metrika*, vol. 67, no. 1, pp. 63–81, 2008, doi: 10.1007/s00184-006-0122-3.
- [16] R Core Team, *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2024