Probability Analysis of Student Attendance Using the Poisson Distribution Model in a Vocational High School

Marsini Sibuea, Muhammad Iqbal

Abstract

Student attendance is essential to the continuity of learning, particularly in vocational high schools (SMK), where students must master both theoretical knowledge and practical skills. This study describes the attendance pattern of Grade XI students in the Television Program Production and Broadcasting (PSPT) program at SMK Negeri 9 Medan using a Poisson distribution approach. A descriptive quantitative design was employed, and data were collected through documentation of attendance records for the 2025/2026 academic year, covering July 2025 to April 2026. All 36 students in the class were included through total sampling. Over 187 effective school days, 6,490 attendances were recorded, yielding a mean of 34.71 students present per day, or approximately 96.41% of the class. Under this parameter, the highest theoretical probability within the range x = 30 to 36 occurred when 34 students were present (6.7759%), whereas the probability of full attendance (36 students) was 6.4774%. However, the Kolmogorov–Smirnov test applied to the 10 monthly means yielded a p-value of 0.0142; therefore, at the 0.05 significance level, the hypothesis that the data follow a Poisson distribution was rejected. Accordingly, the resulting probabilities should be interpreted as model-based estimates rather than as evidence that the Poisson model fits these data. The study underscores the importance of examining model assumptions and daily data before the results are used as a basis for school decision-making.

Keywords: Poisson Distribution; Student Attendance; Probability; Educational Statistics; Vocational High School.

Marsini Sibuea[1]

[1]Magister Teknologi Informasi, Universitas Panca Budi, Indonesia

e-mail: sibueatoriti@gmail.com [1]

Muhammad Iqbal[2]

[2] Magister Teknologi Informasi, Universitas Panca Budi, Indonesia

e-mail : muhammadiqbalpb@gmail.com [2]

2nd International Conference on Islamic Community Studies (ICICS)
Theme: History of Malay Civilisation and Islamic Human Capacity and Halal Hub in the Globalization Era
https://proceeding.pancabudi.ac.id/index.php/ICIE/index

Introduction

Student attendance is one of the determinants of successful learning. Students who attend regularly are more likely to follow the teacher's explanations, take part in discussions, complete exercises, and receive direct feedback. In vocational schools, attendance also determines students' opportunities to take part in practical sessions, which cannot always be replaced by independent study. When students are frequently absent, learning becomes fragmented and the development of their competencies may be hindered [1]–[3].

A number of studies have examined the relationship between attendance and learning outcomes. Sugiharto et al. [1] found a relationship between attendance rates and student learning outcomes. Annisa et al. [4] investigated the association between attendance and academic achievement, while Putri et al. [5] highlighted the relationship of active participation and attendance with competency attainment in vocational high schools. These findings indicate that attendance data are more than an administrative requirement. When properly analyzed, such data can help schools understand how consistently students engage in learning activities.

On the other hand, recording attendance manually or through non-integrated systems often causes problems, such as delayed recapitulation, recording errors, and difficulty in monitoring changes in attendance over time. Digital attendance systems have been developed to address these problems [6], [7]. However, the availability of data does not by itself explain attendance patterns; the data still need to be analyzed with appropriate methods to obtain a more measurable picture.

Statistics offers many models for this purpose. One of them is the Poisson distribution, a discrete probability distribution commonly used to describe the number of events occurring within a given interval of time or space at a certain average rate [8], [9]. In education, this model can be applied to examine the probability of a given number of students being present on a school day. The application of the Poisson distribution to educational data has been discussed previously [8], [10], and the model has also been used in other fields, for example to model the number of disease cases [11].

Nevertheless, the Poisson distribution cannot simply be applied without examining its assumptions. The distribution assumes that the data are counts, that events occur independently within a fixed interval, and that the mean equals the variance. In practice, however, the number of students present in a class is bounded by the number of enrolled students. The fit of the model must therefore be tested, and conclusions cannot rest on the mean alone.

Previous studies have established the importance of attendance, attendance management, and the application of probability models. However, studies that specifically describe the probability of the daily number of students present in a single class at SMK Negeri 9 Medan remain limited. This study therefore focuses on Grade XI PSPT, a class of 36 students observed over ten effective months in the 2025/2026 academic year.

The novelty of this study lies not in developing the Poisson distribution as a new method, but in applying it to interpret class attendance data in a measurable way. The analysis covers the calculation of mean attendance, the estimation of probabilities for specific numbers of students present, and an assessment of model fit. This study does not aim to infer the causes of student absence, since factors such as health, family circumstances, motivation, or transportation constraints are not recorded in the available documentation.

Based on the foregoing, the research questions are: (1) what is the mean number of students present per day during the observation period; (2) what is the probability of a given number of students being present according to the Poisson model; and (3) does the observed attendance pattern conform to the Poisson distribution based on a goodness-of-fit test?

This study aims to calculate the mean daily attendance, determine the probability of specific numbers of students being present, and evaluate the fit of the Poisson model to the data. The results are expected to serve as preliminary input for the school in interpreting class attendance trends and planning attendance monitoring. Nevertheless, this probabilistic information should be used with caution and should not replace attention to students who require individual support.

Literature Review

Student attendance is a record of students' participation in learning activities on the days or hours set by the school. In this study, attendance is treated as quantitative data, namely the number of students recorded as present on each effective school day. School records classify absence into three categories: unexcused absence (alpa), excused absence (izin), and sick leave (sakit). These categories distinguish attendance status, but the main analysis focuses on the number of students present.

The Poisson distribution is a discrete probability distribution used to calculate the probability of x events occurring within an interval when the average rate of occurrence is expressed by the parameter λ. Its probability mass function is given by:

$$P(X = x) = (e^{-}\lambda \times \lambda^{x}) / x!$$

In this formula, $P(X = x)$ is the probability of exactly x events occurring, λ is the mean number of events in the interval, e is Euler's number, and x! is the factorial of x. The value of x is a non-negative integer. In the Poisson distribution, the expected value and the variance are equal, both being λ [8], [9].

The parameter λ can be estimated using Maximum Likelihood Estimation (MLE). For observations $x_1, x_2, ..., x_n$, the MLE of λ is the mean of all observations [12]:

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

In this study, x_i is the number of students present on the i-th effective day, and n is the number of effective days observed. This estimator represents the mean number of students present per day. Its use assumes that each daily observation is recorded with the same definition and time span.

Distributional fit can be evaluated using a goodness-of-fit test. This study employs the Kolmogorov–Smirnov test, which compares the empirical cumulative distribution with the theoretical cumulative distribution. The test statistic is $D = \max|S(x) - F_0(x)|$, where $S(x)$ is the empirical cumulative distribution function and $F_0(x)$ is the cumulative distribution function of the model. The decision is made by comparing the p-value with the significance level $\alpha = 0.05$ [13].

The hypotheses tested are H_0: the data follow a Poisson distribution, and H_1: the data do not follow a Poisson distribution. If the p-value is less than 0.05, H_0 is rejected. If the p-value is equal to or greater than 0.05, there is insufficient evidence to reject H_0. It should be noted that failing to reject H_0 does not prove that the model is absolutely correct.

Research Method

This study adopts a quantitative approach with a descriptive design, as the data analyzed are numerical and were processed using descriptive statistics [14]. The study was conducted at SMK Negeri 9 Medan, with the attendance data of Grade XI PSPT students in the 2025/2026 academic year as its object. The data cover July 2025 to April 2026, or ten effective months of instruction.

The study population comprised all 36 students of Grade XI PSPT. Since all members of the population were accessible, total sampling was applied so that all 36 students were included in the data. The unit of analysis in the distributional calculations is the number of students present on each effective day, not the identity of individual students.

This study uses secondary data obtained from the school's attendance records. The records contain the status present (H), unexcused absence (A), excused absence (I), and sick (S) for each student. The number of students present per day was calculated as the number of enrolled students (36) minus the number of students absent on that day.

The main research variable is the number of students present (X) on each effective day. Since the class consists of 36 students, X takes integer values from 0 to 36. The analysis was carried out in R through several stages: checking data completeness, summarizing the number of students present, estimating the parameter λ using MLE, computing Poisson probabilities for selected values of x, and testing model fit.

Probabilities were computed using the formula $P(X = x) = e^{-\lambda} \lambda^x/x!$. The estimated λ and the values of x representing the number of students present were substituted into the formula, and the results are presented as probabilities and percentages. The resulting values are theoretical probabilities under the model, not frequencies observed directly in the field.

Model fit was assessed using the Kolmogorov–Smirnov test at the 5% significance level. The null hypothesis states that the data follow a Poisson distribution, whereas the alternative hypothesis states otherwise. The decision is made by comparing the p-value with $\alpha = 0.05$. Because λ is estimated from the same data, the standard Kolmogorov–Smirnov procedure is, strictly speaking, not fully appropriate; a valid p-value may require calibration, for instance through parametric simulation. Since the data were available as monthly summaries, the test was applied to the 10 monthly mean attendance values ($n = 10$) rather than to the 187 daily observations.

The Poisson distribution also requires the mean and the variance to be equal, so both must be computed from the same daily data. Monthly summaries are insufficient for calculating the daily variance; this check therefore requires the raw daily data and should be performed before the model is declared appropriate or inappropriate.

The overall research flow is summarized in the conceptual framework shown in Figure 1. Attendance data serve as the input, which is then processed through parameter estimation, probability calculation, and a goodness-of-fit test. The test results are subsequently evaluated against the assumptions of the Poisson distribution before being used as a basis for attendance monitoring by the school.

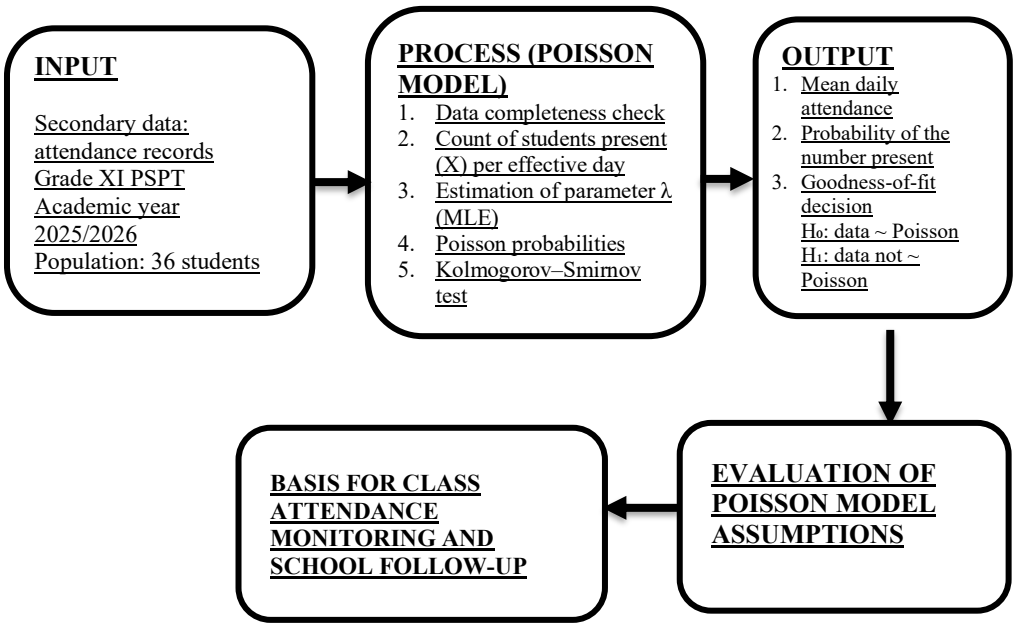


Figure 1. Conceptual Framework of the Study

Results And Discussion

4.1 Overview of Attendance Data

The research data cover ten months of instruction, from July 2025 to April 2026. Based on the attendance summary, there were 187 effective days with a total of 6,490 student attendance records. With 36 students in the class, the maximum possible attendance during the observation period was 187 × 36 = 6,732. The difference between the two, 242, represents the total number of student absences. Monthly details are presented in Table 1.

Table 1. Monthly Student Attendance Data

No.	Month	Semester	School Days	Absent	Total Present	Mean/Day
1	July 2025	First	16	15	561	35.06
2	August 2025	First	21	22	734	34.95
3	September 2025	First	21	23	733	34.90
4	October 2025	First	23	46	782	34.00
5	November 2025	First	20	21	699	34.95
6	December 2025	First	15	16	524	34.93
7	January 2026	Second	19	38	646	34.00
8	February 2026	Second	16	17	559	34.94
9	March 2026	Second	15	19	521	34.73
10	April 2026	Second	21	25	731	34.81
	Total		187	242	6,490	34.71

Source: attendance summary of Grade XI PSPT, SMK Negeri 9 Medan, 2025/2026 academic year.

The mean number of students present per day in each month is visualized in Figure 2. October 2025 and January 2026 fell below the overall mean, whereas the other months remained relatively stable, ranging from 34.73 to 35.06 students per day.

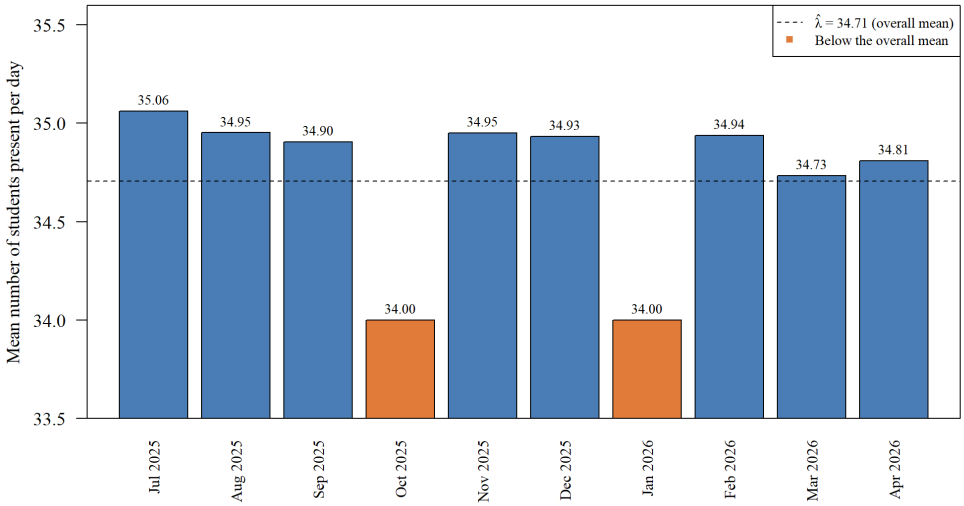


Figure 2. Mean Number of Students Present per Day by Month

4.2 Estimation of Mean Attendance

During the observation period, 6,490 attendances were recorded over 187 effective days. Using the MLE, the mean number of students present per day was calculated as follows:

$$\hat{\lambda} = 6{,}490 / 187 = 34.7059 \approx 34.71 \text{ students per day}$$

This means that, on average, about 34 to 35 of the 36 enrolled students were present in class each day. The mean attendance rate was 34.7059/36 × 100% = 96.41%, while the mean absence rate was 3.59% (242 out of 6,732 possible attendances). These figures indicate that class attendance was relatively high throughout the observation period. However, the class mean does not reveal whether absences were concentrated among particular students or spread across many students.

4.3 Probabilities Based on the Poisson Model

With $\hat{\lambda} = 34.71$, the theoretical probabilities for specific numbers of students present were calculated using the Poisson probability mass function. The results for x = 30 to x = 36 are presented in Table 2.

Table 2. Probabilities Based on the Poisson Model

Number present (x)	P(X = x)	Percentage	Remarks
30	0.051983	5.1983%	
31	0.058197	5.8197%	
32	0.063118	6.3118%	
33	0.066381	6.6381%	
34	0.067759	6.7759%	Highest probability in the table range
35	0.067190	6.7190%	
36	0.064774	6.4774%	Full attendance

Table 2 shows that the highest probability within the displayed range occurs at x = 34, namely 0.067759 or 6.7759%. In other words, under the Poisson model with $\lambda = 34.71$, the theoretical probability of exactly 34 students being present is approximately 6.78%, while the probability of all 36 students being present is 0.064774 or 6.4774%. It should be noted that the Poisson model has no upper bound. The model even assigns a probability of approximately 37.07% to more than 36 students being present, which is impossible because the class has only 36 students. This is one indication that the Poisson distribution is not well suited to attendance data bounded by the number of enrolled students.

These probabilities should be distinguished from the proportions of events actually recorded. A value of 6.7759% does not mean that 34 students were present on 6.7759% of all days; it is a model-based result. To assess how well the model describes the data, the probabilities must be interpreted together with the goodness-of-fit test and the assumption checks.

Figure 3 compares the probability distribution of the Poisson model with a binomial distribution (n = 36; p = 0.9641) that has the same mean of 34.71. The Poisson distribution is considerably wider, and part of its probability mass lies above 36 students. In contrast, the binomial distribution is concentrated between 34 and 36 students and never exceeds the number of enrolled students.

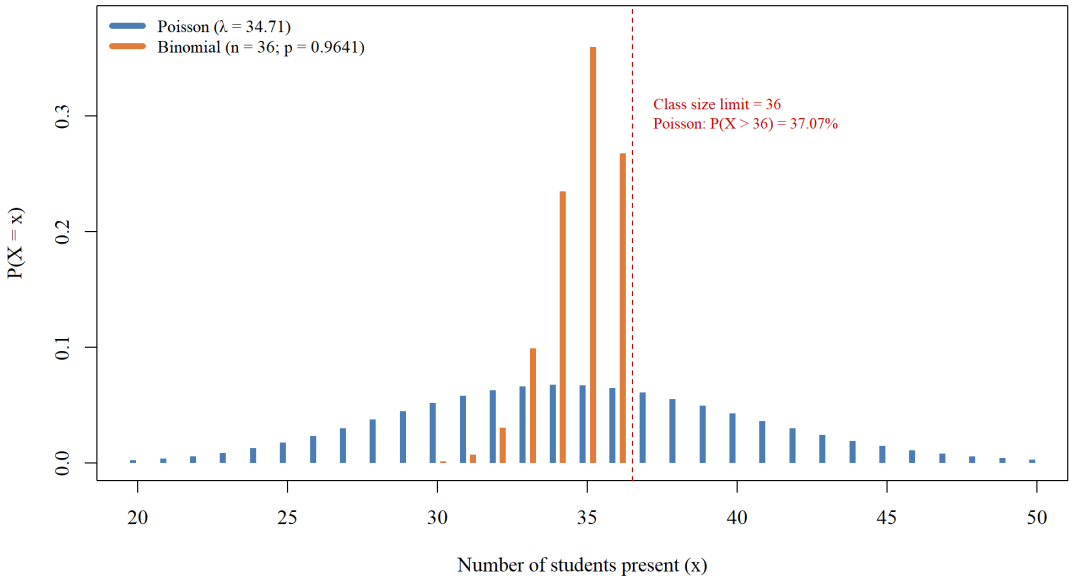


Figure 3. Probability Distributions of the Poisson and Binomial Models

4.4 Goodness-of-Fit Test Results

The Kolmogorov–Smirnov test on the 10 monthly means (n = 10) yielded a D statistic of 0.4974 with a p-value of 0.0142. At the significance level $\alpha = 0.05$, this p-value is smaller than α. With the hypotheses H_0: the data follow a Poisson distribution and H_1: the data do not follow a Poisson distribution, the decision is to reject H_0.

Table 3. Goodness-of-Fit Test Results

Test component	Value
Kolmogorov–Smirnov statistic (D)	0.4974
p-value	0.0142
Number of observations (n)	10 (monthly means)
Significance level (α)	0.05
Statistical decision	H_0 rejected

Accordingly, the test results do not support the conclusion that the attendance data follow a Poisson distribution at the 5% level.

A comparison between the empirical cumulative distribution function and the Poisson cumulative distribution function shows that the largest discrepancy occurs at x = 34: the Poisson model already assigns a cumulative probability of 0.4974, whereas none of the monthly means falls below 34. This discrepancy constitutes the D statistic. The comparison also shows that the empirical data are tightly clustered between 34 and 35, whereas the Poisson distribution is spread much more widely.

4.5 Discussion and Implications

The mean attendance rate of 96.41% indicates that most students were recorded as present during the observation period. Nevertheless, the monthly data reveal some variation. The lowest mean attendance occurred in October 2025 and January 2026, at 34.00 students per day each, consistent with the highest numbers of absences in those two months (46 and 38 records, respectively). Conversely, the highest mean was recorded in July 2025, at 35.06 students per day. These differences may serve as an initial signal for the school to review the

attendance records for those months. Because this study did not examine the causes of these changes, the underlying factors cannot be determined from the available data.

This pattern is illustrated in Figure 4. The absence rates in October 2025 and January 2026 were each 5.56%, approximately 1.5 times the mean absence rate over the observation period (3.59%). In the remaining months, absence rates ranged from 2.60% to 3.52%.

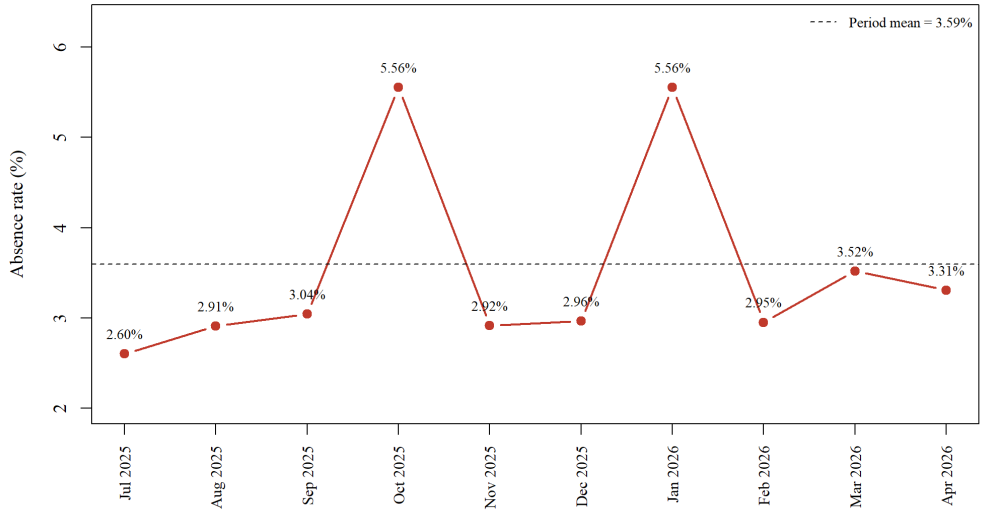


Figure 4. Monthly Student Absence Rate

This finding is consistent with the importance of attendance management discussed in studies on the relationship between attendance and learning outcomes [1], [4], [5]. However, this study did not measure the effect of attendance on academic performance. Its main contribution is a quantitative summary and an example of applying probability to class attendance data. For decision-making, the school can use this summary as a basis for monitoring and follow it up by reviewing individual student records.

4.6 Limitations of the Study

This study has several limitations. First, the research object was a single class of 36 students, so the results cannot be directly generalized to all students of SMK Negeri 9 Medan. Second, the data were derived from attendance summaries over ten effective months and therefore depend heavily on the accuracy of the school's administrative records. Third, this study did not investigate the reasons for absence or relate them to academic achievement. Fourth, the goodness-of-fit test was performed on 10 monthly means rather than on daily data, and the results did not support the fit of the Poisson distribution, as the p-value was below 0.05. The daily data and the testing procedure should be re-examined before these results are used to assess the suitability of the Poisson model.

Future studies may involve more classes or grade levels, use verified daily data, and compare the Poisson distribution with other, more suitable count-data models, such as the binomial distribution, which has a natural upper bound equal to the number of students. Further analyses could also examine attendance patterns by student, by day of the week, or by semester to obtain a more complete picture, without presupposing the causes of absence from the outset.

Conclusion

Based on the attendance summary of Grade XI PSPT students at SMK Negeri 9 Medan from July 2025 to April 2026, there were 187 effective days with 6,490 attendance records. The mean number of students present, estimated using Maximum Likelihood Estimation, was 34.71 out of 36 students per day, or approximately 96.41%. Under this parameter, the highest theoretical probability within the range x = 30 to 36 occurred when 34 students were present (6.7759%), whereas the probability of all 36 students being present was 6.4774%.

The Kolmogorov–Smirnov test on the 10 monthly means yielded D = 0.4974 and a p-value of 0.0142. Since the p-value was less than 0.05, the hypothesis that the data follow a Poisson distribution was rejected. In other words, this study cannot conclude that the Poisson distribution is appropriate for these attendance data. The calculated probabilities may still be presented as theoretical estimates based on the Poisson parameter, provided that the model fit is re-verified using daily data and an appropriate testing procedure.

In practical terms, the attendance summary helps the school monitor month-to-month changes in attendance rates and identify periods that warrant further review. Future research is recommended to examine the variance of the daily data, validate the goodness-of-fit results, and compare alternative statistical models, such as the binomial distribution, when the Poisson assumptions are not met.

References

[1] R. R. A. Sugiharto, K. L. P. Gunawan, M. Wulandari, and A. Wulandari, “Hubungan Kehadiran Siswa dengan Hasil Belajar Siswa Kelas XII di SMA Hang Tuah Surabaya,” Jurnal Pendidikan Tambusai, vol. 9, no. 3, pp. 36405–36413, 2025.

[2] E. Murniati, D. P. Sari, and T. Hidayat, “Dampak Ketidakhadiran Siswa terhadap Kontinuitas Proses Belajar Mengajar di Sekolah Menengah Kejuruan,” Jurnal Pendidikan Vokasi, vol. 13, no. 1, pp. 21–33, 2023.

[3] L. Naibaho, M. Hutabarat, and P. Siagian, “Pengaruh Frekuensi Kehadiran terhadap Pemahaman Materi Siswa Berdasarkan Teori Behavioristik,” Jurnal Penelitian Pendidikan, vol. 10, no. 3, pp. 112–122, 2023.

[4] N. Annisa, A. Maulana, and R. Fitri, “Pengaruh Tingkat Kehadiran terhadap Prestasi Belajar Siswa SMA di Banda Aceh,” Jurnal Pendidikan dan Pembelajaran, vol. 12, no. 2, pp. 45–54, 2023.

[5] A. M. Putri, N. L. Sari, and D. Rahmawati, “Hubungan Partisipasi Aktif dan Kehadiran Siswa terhadap Capaian Kompetensi di SMK,” Jurnal Pendidikan Kejuruan Indonesia, vol. 5, no. 2, pp. 67–78, 2025.

[6] A. J. Budiarti and J. Aryanto, “Pengembangan Aplikasi Sistem Presensi Siswa dengan Integrasi Fitur Kehadiran Per Mata Pelajaran dan Izin Digital,” Bulletin of Computer Science Research, vol. 6, no. 1, pp. 275–283, 2025.

[7] Y. Pratama, B. Santoso, and H. Nugroho, “Sistem Informasi Absensi Berbasis Digital dan Dampaknya terhadap Kualitas Pengelolaan Pembelajaran di Sekolah Menengah,” Jurnal Teknologi Informasi dan Pendidikan, vol. 17, no. 1, pp. 22–31, 2024.

[8] S. Muharani, D. Rahmi, A. Kurniati, and S. Yuniati, “Distribusi Poisson: Analisa Data Lama Masa Studi Mahasiswa UIN Suska Riau,” RELEVAN: Jurnal Pendidikan Matematika, vol. 4, no. 6, pp. 2649–2658, 2024.

[9] R. Septiani, Y. Yundari, and E. Sulistianingsih, “Distribusi Probabilitas Diskrit: Kajian Teoritis dan Aplikasinya dalam Pemodelan Statistika,” Jurnal Matematika Integratif, vol. 18, no. 2, pp. 95–106, 2022.

[10] D. Sulistiawati and R. Anggraini, “Penerapan Distribusi Poisson dalam Pemodelan Kejadian Diskrit pada Data Pendidikan,” Jurnal Ilmiah Matematika dan Pendidikan Matematika, vol. 15, no. 1, pp. 44–53, 2023.

[11] F. R. Firdaus, S. M. Ismauz, and A. H. Sianipar, “Model Distribusi Poisson dan Gamma untuk Analisis Penyakit TBC di Daerah Provinsi Jawa Barat,” Cendekia: Jurnal Penelitian dan Pengkajian Ilmiah, vol. 2, no. 1, pp. 117–125, 2025.

[12] H. Prasetyo, N. Fitriani, and A. Wahyudi, “Estimasi Parameter Distribusi Poisson Menggunakan Metode Maximum Likelihood Estimation,” Jurnal Statistika dan Komputasi, vol. 5, no. 1, pp. 11–20, 2023.

[13] M. A. Rahman and I. Setiawan, “Penerapan Uji Kecocokan Kolmogorov–Smirnov pada Pemodelan Distribusi Data Pendidikan,” Jurnal Analisis Statistika, vol. 9, no. 2, pp. 73–84, 2022.

[14] I. P. Sari, A. S. Nugroho, and W. Lestari, “Penerapan Metode Statistika Deskriptif dalam Analisis Data Akademik Siswa Sekolah Menengah,” Jurnal Statistika dan Aplikasinya, vol. 8, no. 1, pp. 33–44, 2024.