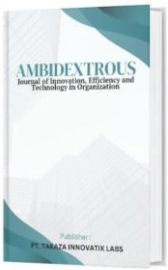
Ambidextrous: Journal of Innovation, Efficiency and Technology in Organization
https://journal.takaza.id/index.php/ambidextrous
Vol. 4 , No. 1 , 2026, pp. 18-26
E-ISSN: 3031-7002
E-mail:ambidex@takaza.id



Comparison of Poisson Inverse Gaussian (PIG) and Negative Binomial (NBR) Regression Models Using Generalized Linear Models (GLM) in Dengue Fever Cases in East Java Province in 2024

NindyAulia Eka Puspita
Mathematics Study Program, Indonesian Defense University, Surabaya
Corresponding Author e-mail:nindyaulia.eka@gmail.com

Article History:
Received: 15-01-2026
Revised: 03-02-2026
Accepted: 07-02-2026

Keywords: Akaike Information Criterion, DBD Model, Negative Binomial Regression, Overdispersion, Poisson Inverse Gaussian

Abstract: Background: *Dengue Hemorrhagic Fever (DHF) is a major health problem in East Java Province due to population density, sanitation, availability of medical personnel, and altitude. DHF count data often experiences overdispersion, challenging the Poisson regression assumption in the Generalized Linear Model (GLM).* ***Objective:*** *To compare the performance of Poisson, Negative Binomial (NBR), and Poisson Inverse Gaussian (PIG) models in modeling DHF cases.* ***Method:*** *A cross-sectional quantitative study analyzed 38 districts/cities in East Java (saturation sampling) using secondary data processed with .* ***Analysis:*** *Descriptive statistics, multicollinearity test (VIF<10), overdispersion test, and model comparison via AIC/deviance.* ***Results:*** *NBR achieved the lowest AIC (412.10) compared to Poisson (1466.60) and PIG (412.41), confirming significant overdispersion (deviance/df ≈37.16). Population density and altitude had a significant effect (p<0.05).* ***Conclusion:*** *NBR is optimal for DHF modeling, supporting targeted environmental and health interventions in East Java.*

How to Cite: Nindy Aulia Eka Puspita (2026). Comparison of Poisson Inverse Gaussian (PIG) and Negative Binomial (NBR) Regression Models Using Generalized Linear Models (GLM) in Dengue Fever Cases in East Java Province in 2024. *Ambidextrous Journal of Innovation Efficiency and Technology in Organizations. 4*(1). 18-26. https://doi.org/10.61536/ambidextrous.v4i1.414

https://doi.org/10.61536/ambidextrous.v4i1.414

Introduction

Dengue Hemorrhagic Fever (DHF) remains a major public health problem in Indonesia, particularly in East Java Province, which has recorded high cases due to population density and environmental conditions that support the breeding of the Aedes aegypti mosquito. Ministry of Health data shows an endemic pattern with an annual increase in cases, influenced by factors such as sanitation, rainfall, and the availability of medical personnel (Ministry of Health, 2016). In 2024, East Java was among the provinces with significant incidents, contributing thousands of national cases amid a national upward trend.

Dengue fever case data are count data, suitable for modeling with Poisson regression in the Generalized Linear Model (GLM), but often experience overdispersion where the variance exceeds the mean, causing biased parameter estimates and unreliable standard errors (Eminita et al., 2019). This overdispersion is common in dengue fever epidemiological data in Indonesia, requiring alternatives such as Negative Binomial Regression (NBR) which adds a dispersion parameter for variances greater than the mean (Suryani et al., 2021).

Poisson Inverse Gaussian (PIG) regression has emerged as a more flexible option for extreme overdispersion and heterogeneity, with a wider range of skewness than NBR, making it more adaptable to complex, inter-regional dengue fever data (Suleman & Indriyani, 2022). Comparative studies have shown that PIG often outperforms long-tailed data, but direct comparisons with NBR and Poisson in the East Java context are still limited (Guntur & Da Rato, 2024). This raises the need for systematic evaluation using criteria such as AIC and deviance to determine the optimal model.

This study compares the performance of Poisson, NBR, and PIG models on 2024 East Java dengue case data with population density, medical personnel, proper sanitation, and altitude variables, to address overdispersion and detect significant factors (Hossain et al., 2023). The urgency lies in supporting dengue control policies in endemic areas with the latest data, while the novelty of the study lies in the comprehensive analysis of PIG versus NBR in East Java, going beyond previous studies that were limited spatially or temporally.

Research Methods

Types and Methods of Research

This study uses a quantitative approach with a non-experimental cross-sectional design to analyze the relationship between independent variables and the number of Dengue Hemorrhagic Fever (DHF) cases in East Java Province in 2024. The Generalized Linear Model (GLM) method is applied as the main framework for modeling enumerated data, including Poisson, Negative Binomial (NBR), and Poisson Inverse Gaussian (PIG) regression to address overdispersion common in epidemiological data. This approach is suitable for aggregated secondary data per district/city, allowing for comparative evaluation of models based on criteria such as AIC and deviance (Cresswell & Creswell, 2017; Sugiyono, 2013; Eminita et al., 2019).

Data Analysis Instruments and Techniques

The research instrument consisted of secondary data in the form of the number of DHF cases as the response variable (Y) and four predictors, namely population density (X1), number of medical personnel (X2), proper sanitation (X3), and area altitude (X4), which were processed using R software with packages such as gamlss for PIG. Analysis techniques included descriptive statistics (min, max, mean, SD), multicollinearity tests via Variance Inflation Factor (VIF < 10), overdispersion tests (deviance ratio/df > 1), and model comparisons using AIC, deviance, and log-likelihood. Spatial validation was performed with Moran's I on residuals to detect autocorrelation (Sudaryono, 2022; Emzir, 2021; Suryani et al., 2021).

Population and Sample

The study population included 38 districts/cities in East Java Province in 2024, with the total sample using saturation sampling (census) techniques to ensure complete representation of cross-sectional data from official sources such as the East Java Health Office, BPS, BIG, and BMKG. This sample selection avoided bias because the data was public aggregate, covering a variety of regions from densely populated areas such as Surabaya to rural eastern provinces (Cresswell & Creswell, 2017; Sugiyono, 2013).

Research Procedures

The procedure begins with secondary data collection from official sources, followed by descriptive analysis and spatial visualization using R for case distribution maps. Next, multicollinearity checks (correlation and VIF), stepwise Poisson-NBR-PIG modeling, overdispersion tests, simultaneous (Likelihood Ratio Test) and partial (z-test, $p < 0.05$) hypothesis testing, and selection of the best model via the smallest AIC are performed. This process concludes with interpretation of significant factors and validation of spatial residuals, ensuring a systematic flow from exploration to conclusion (Emzir, 2021; Sudaryono, 2022; Guntur & Da Rato, 2024).

Results and Discussion

Characteristics of the Number of Dengue Fever (DHF) Cases in East Java in 2022

An initial analysis was conducted by depicting the spatial distribution of Dengue Hemorrhagic Fever (DHF) cases in East Java Province in 2024 to determine the distribution pattern of cases in each district/city. This map visualization, using R Studio, provides an initial overview of differences in DHF case rates between regions and supports further analysis using statistical methods.

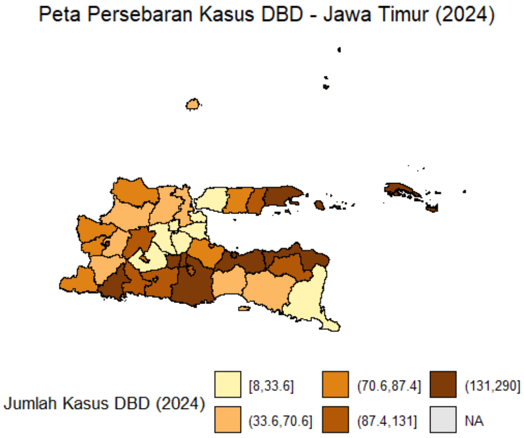


Figure 1.Map of the Distribution of Dengue Fever Cases in East Java Province in 2020

Figure 1 shows that dengue fever cases are unevenly distributed across East Java. Areas with high case numbers are generally densely populated, such as Surabaya, Sidoarjo, and Malang. Meanwhile, areas with low case numbers are in the eastern and western parts of the province. Therefore, this pattern suggests the influence of population density and environmental conditions on the spread of dengue fever.

Descriptive Analysis

Descriptive analysis was conducted to provide a general overview of the characteristics of the research data. Descriptive statistics included the minimum, maximum, average (mean), and standard deviation values of each variable used in the regression model analysis.

Table 1. Descriptive statistics of research variables

Variables	Min	Mean	Max	StDev
Dengue Hemorrhagic Fever	8.00	90.71	290.00	64.09153
Population density	327.00	2,632.00	6,987.00	3.792226
Number of Medical Personnel	2,063	5,682	19,359	160,554.30
Proper Sanitation	29.49	88.21	101.38	13.08836
Area Altitude	2.00	81.13	556.00	127,257

https://doi.org/10.61536/ambidextrous.v4i1.414

Multicollinearity Examination

A multicollinearity test is performed to ensure there are no strong relationships between independent variables that could affect the stability of the regression model. This test uses correlation coefficient analysis between the independent variables.

Table 2. Correlation Coefficient between Variables

Variables		Y	X_1	X_2	X_3	X_4
Y	Coefficient Correlation	1	-0.37	0.33	0.11	0.26
	p-value		0.022	0.044	0.53	0.115
X_1	Coefficient Correlation	-0.37	1	-0.55	-0.14	0.11
	p-value	0.022		<0.001	0.399	0.525
X_2	Coefficient Correlation	0.33	-0.55	1	0.35	0.11
	p-value	0.044	<0.001		0.029	0.507
X_3	Coefficient Correlation	0.11	-0.14	0.35	1	-0.06
	p-value	0.53	0.399	0.029		0.722
X_4	Coefficient Correlation	0.26	0.11	0.11	-0.06	1
	p-value	0.115	0.525	0.507	0.722	

Based on Table 2, all correlation coefficients for each independent variable are below 0.80, thus concluding that there is no multicollinearity between the variables. This indicates that each independent variable in the model has a sufficiently independent relationship with one another and is suitable for use in regression analysis.

Table 3. VIF Values of Predictor Variables

Variables	VIF
X_1	1.506597
X_2	1.706848
X_3	1.167191
X_4	1.071755

Based on Table 3, all Variance Inflation Factor (VIF) values for the predictor variables are less than 10, ranging from 1.07 to 1.70. These values indicate no signs of multicollinearity among the independent variables. Therefore, all predictor variables are considered suitable for use in further regression modeling.

Poisson Regression

The Poisson Regression Model was used because the number of Dengue Hemorrhagic Fever (DHF) cases is count data, which describes the number of disease incidents in each district/city. Through this model, parameter estimation was performed to determine the direction and magnitude of the influence of independent variables, including Population Density, Number of Medical Personnel, Adequate Sanitation, and Area Altitude, on the number of DHF cases. The results of the Poisson Regression model parameter estimation are shown in Table 4 below.

Table 4.Parameter Estimation of Poisson Regression Model

Parameter	Estimate	Std. Error	t-count (z)	p-value
(Intercept)	4.6170	0.1404	32,878	< 2e-16 ***
Population density	-0.0001660	0.00001485	-11,178	< 2e-16 ***
Proper Sanitation	0.001401	0.001527	0.918	0.359
Number of Medical Personnel	0.007485	0.005680	1,318	0.188
Area Altitude	0.001342	0.000121	11,122	< 2e-16 ***
Null Deviance= 1604.3, Df = 37 **Residual Deviance**= 1226.3, Df = 33 **AIC**= 1466.6				

Based on Table 4, it is known that the Population Density and Altitude variables significantly influence the number of dengue fever cases (p-value <0.05). Meanwhile, the Proper Sanitation and Number of Medical Personnel variables do not. The AIC value of 1466.6 and the Residual Deviance of 1226.3 indicate that the Poisson model is able to describe the data quite well. However, further testing is needed to ensure there is no overdispersion in the data.

Overdispersion Test

The overdispersion test is performed to determine whether the Poisson Regression model meets the basic assumption that the variance is equal to the expected value. If the variance of the data is greater than the mean, overdispersion occurs, which causes the Poisson model to no longer be appropriate and produces inefficient parameter estimates.$(Var(Y_i) = \mu_i)$

The overdispersion test hypothesis used is as follows:

a) $H_0: Var(Y) = \mu_i$ (There is no overdispersion, according to the Poisson assumption)

b) $H_1: Var(Y) = \mu_i + \alpha g(\mu_i)$(Overdispersion occurs,$\alpha > 0$

Based on the summary results of the Poisson model, the residual deviance value is 1226.3 with residual degrees of freedom = 33. The dispersion measure value is calculated as the ratio of deviance to degrees of freedom, namely:

$$\widehat{\emptyset} = \frac{Residual\ Deviance}{df\ residual} = \frac{1226.3}{33} \approx 37.16$$

To test the significance of overdispersion, the chi-square test is used on the residual deviance: $p - value\ = 1 - F_{x^2_{33}}(1226.3)$.These values are very small (almost zero; , so with = $0.05.pp < 0.001p < aa$

Since the -value is <0.05, the test is rejected. Thus, there is strong evidence that the data is overdispersed. Therefore, the Poisson model is inadequate, and an alternative model that can accommodate overdispersal (e.g., Negative Binomial Regression or Inverse Gaussian Poisson Regression) should be used.$\widehat{\emptyset} \approx 37.16 > 1pH_0$

Inverse Gaussian Poisson Regression Model

Based on the results of the overdispersion test which showed that the Poisson Regression model was not suitable for use because the data variance was greater than the average value, as well as the results of the Negative Binomial Regression analysis which had corrected most of the overdispersion problems, the analysis was continued using Poisson Inverse Gaussian (PIG) Regression. This model is a development of the Poisson model which is able to accommodate more extreme levels of overdispersion by adding a dispersion parameter (τ), thus providing more efficient and reliable parameter estimation results.

https://doi.org/10.61536/ambidextrous.v4i1.414

Inverse Gaussian Poisson regression modeling was performed using the gamlss package in R software, using the Maximum Likelihood Estimation (MLE) method. This approach was chosen because it produces stable parameter estimates and allows adjustments to inhomogeneous data distributions.

In general, the functional form of the Inverse Gaussian Poisson Regression model in this study can be written as follows:

$$\mu_1 = \exp\,(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4)$$
$$\mu_2 = \exp\,(\beta_0 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4)$$
$$\mu_3 = \exp\,(\beta_0 + \beta_3 X_3 + \beta_4 X_4)$$
$$\mu_4 = \exp\,(\beta_0 + \beta_4 X_4)$$
$$\mu_5 = \exp\,(\beta_0 + \tau)$$

The four models were successfully converged and used to identify the most optimal combination of independent variables in explaining the variation in the number of Dengue Hemorrhagic Fever (DHF) cases in East Java Province in 2024. The parameter estimation results of the four models are presented in Table 6 below.

Table 5. Parameter Estimates of Inverse Gaussian Poisson Regression

Variables	Model 1	Model 2	Model 3	Model 4
β_0 (Intercept)	4,893	5,014	5.126	5,182
β_1 (Population Density)	-0.00019	-0.00018	-0.00017	-
β_2 (Number of Medical Personnel)	0.00721	0.00697	-	-
β_3 (Proper Sanitation)	0.00152	-	-	-
β_4 (Area Altitude)	0.00136	0.00133	0.00132	0.00128
τ (Dispersion impact)	2,581	2,479	2,365	2,142

Based on Table 5, all models show convergent parameter estimates with relatively consistent direction of variable influence. The Population Density variable has a negative effect on the number of dengue cases, indicating that increasing population density is not always followed by an increase in the number of cases, possibly due to differences in urbanization levels and sanitation quality between regions. Conversely, the Altitude variable shows a positive effect on dengue cases, which may be attributed to variations in climate and mosquito habitats at certain altitudes.

The relatively small dispersion parameter () values across all models indicate that the Inverse Gaussian Poisson Regression model is able to adapt well to data variations, so this model has the potential to provide more efficient modeling results than the Poisson and Negative Binomial models. Furthermore, the selection of the best model will be made based on a comparison of the Akaike Information Criterion (AIC) values.τ

Hypothesis Testing

In hypothesis testing there are two tests, namely simultaneous tests and partial tests to determine which parameters are significant.

1. **Simultaneous Testing**

   This test is conducted simultaneously, so it is likely that this test fits the Poisson Inverse Gaussian Regression model. The following is the test hypothesis:

   $H_0 : \beta_1 = \beta_2 = \cdots \beta_k$(independent variables do not affect the independent variables).

   H_1 : there is at least one (independent variable that influences the dependent variable)$\beta_j \neq j = 1,2,\ldots,k$

   Simultaneous testing is carried out using the G Test Statistics (Likelihood Ratio Test) which is defined as follows:

$$G = 2(L_1 - L_0)$$

where is the log-likelihood of the model with the independent variable, and is the log-likelihood of the model without the independent variable. The test statistic G follows a distribution with as many degrees of freedom as the number of parameters and the significance level .$L_1 L_0 X^2 (v = k) a = 0{,}05$
Decision making criteria:

a) $G > X^2_{(a,v)}$then reject.H_0

b) $G \leq X^2_{(a,v)}$then accept.H_0

Table 6. Parameter Estimates of Inverse Gaussian Poisson Regression

Model	**G Statistics**	**v**	$X^2_{(0.05,V)}$	**Decision**
Model 1	625.91	7	14.07	RejectH_0
Model 2	626.54	8	15.50	RejectH_0
Model 3	631.41	9	16.92	RejectH_0
Model 4	648.10	10	18.30	RejectH_0

In table 6 above, it can be seen that there is rejection because the test statistic value of G in each model is greater than, so it can be concluded that the parameters have a significant effect on the model.$H_0 G > X^2_{(a,v)}$

2. Partial Testing

Partial or individual parameter testing can be used to determine which independent variables have a significant influence on the dependent variable by looking at the rejection criteria if
$|Z_{hit}| > Z_{\frac{a}{2}} = 1{,}96$ or p-value < where is the significance level.aa
The following is the test hypothesis:
$H_0 : \beta_j = 0, j = 1,2, \dots, k$(independent variables do not affect the dependent variable).
$H_1 : \beta \neq 0, j = 1,2, \dots, k$(independent variables influence dependent variables)
decision making criteria:

a) $|Z_{hit}| > Z_{\frac{a}{2}} = 1.96$or p-value < 0.05 then it is rejectedH_0

b) $|Z_{hit}| > 1{,}96$or p-value 0.05≥

Table 7. Partial testing

Model	**Significant Parameters**
Model 1	$\beta_1, \beta_2, \beta_4$
Model 2	β_1, β_4
Model 3	β_1, β_4
Model 4	β_1, β_4

In table 7 above, it can be seen that in all models there are significant parameters.The Population Density (β_1) and Area Altitude (β_4) variables consistently showed values and p-values <0.05, meaning they significantly influenced the number of dengue fever cases. Meanwhile, the Proper Sanitation (β_3) and Number of Medical Personnel (β_2) variables did not have a significant effect because they had p-values >0.05.$|Z_{hit}| > 1{,}96$

Best Model Selection of Inverse Gaussian Poisson Regression

The best model for Poisson Inverse Gaussian (PIG) regression is selected by comparing the Akaike Information Criterion (AIC) values of several converged models. The model with the smallest AIC value is considered the most efficient in explaining the data without adding excessive parameter complexity. The AIC values for each model are presented in Table 8 below.

Table 8. AIC of the Inverse Gaussian Poisson model

Model	AIC	Log-Likelihood	df
Model 1	412.4068	-201.2034	4
Model 2	413.2838	-202.6419	3
Model 3	414.3685	-201.1843	5
Model 4	415.0738	-202.5369	4

Negative Binomial Regression

After it was discovered that the Poisson model was overdispersive, the analysis continued using a Negative Binomial Regression model. This model can accommodate data variance greater than the mean, providing more stable and accurate parameter estimates. The results of the model parameter estimates are presented in Table 9 below.

Table 9. Parameter Estimation of Negative Binomial Regression Model

Parameter	Estimate	Std. Error	t-count (z)	p-value
(Intercept)	5.0050	0.7703	6,497	8.17e-11 ***
Population density	-0.0002449	0.00007939	-3,084	0.00204 **
Proper Sanitation	-0.00004755	0.00008645	-0.006	0.99561
Number of Medical Personnel	-0.01267	0.03613	-0.351	0.72588
Area Altitude	0.001933	0.0008488	2,277	0.02279 *
Null Deviance= 52.955, Df = 37 **Residual Deviance**= 40.447, Df = 33 **AIC**= 412.1				

Based on Table 9, the Population Density and Area Altitude variables significantly influence the number of DHF cases (p-value <0.05). The other variables, namely Proper Sanitation and Number of Medical Personnel, did not significantly influence the number of cases. The AIC value of 412.1 indicates that the Negative Binomial Regression model provides a better fit than the previous Poisson model.

Comparison of AIC Values of Three Regression Models

The best model was selected by comparing the Akaike Information Criterion (AIC) values of the three regression models used: Poisson Regression, Negative Binomial Regression, and Inverse Gaussian Poisson Regression. The model with the smallest AIC value was considered to have the best fit to the data with efficient parameter complexity. The AIC values for each model are presented in Table 10 below.

Table 10. AIC of three regression models

Regression Model	Log-Likehood	AIC
Poisson	-733.30	1466.60

https://doi.org/10.61536/ambidextrous.v4i1.414

Poisson Inverse Gaussian	-206.05	412.10
Negative Binomial	-201.20	412.41

Based on Table 10, the Negative Binomial Regression model has the smallest AIC value, at 412.10, compared to the Inverse Gaussian Poisson (412.41) and Poisson (1466.60) models. This indicates that the Negative Binomial Regression model provides the most appropriate modeling results in describing the variation in the number of dengue fever cases in East Java Province in 2024. Therefore, this model was selected as the best model in this study.

Conclusion

This study concludes that the Negative Binomial Regression (NBR) model is the best approach to modeling the number of Dengue Hemorrhagic Fever (DHF) cases in East Java Province in 2024, with the smallest Akaike Information Criterion (AIC) value of 412.10 compared to the Poisson model (1466.60) and Poisson Inverse Gaussian (412.41), supported by a significant overdispersion test (deviance ratio/df$\approx$37.16, $p<0.001$). Population density and regional altitude variables have a significant effect on the increase in cases ($p<0.05$), while proper sanitation and the number of medical personnel are not significant, confirming the overdispersion of the enumeration data due to regional heterogeneity.

However, limitations of the study include the use of aggregate secondary data per district/city that limits micro-spatial analysis, as well as the non-spatial GLM approach that does not yet accommodate full geographic autocorrelation, as initially detected via Moran's I. For further research, it is recommended to integrate spatial models such as Geographically Weighted Negative Binomial Regression (GWNB) or INLA with the latest monthly data, as well as dynamic variables such as BMKG rainfall. Practically, these findings support the policy of prioritizing sanitation equity in dense lowland areas and optimizing medical personnel, contributing to the control of dengue fever endemic in East Java.

References

Cresswell, J. W., & Creswell, J. D. (2017). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Sage Publications.

Emzir. (2021). Quantitative research methodology. Pustaka Setia.

Eminita, V., Kurnia, A., & Sadik, K. (2019). Handling overdispersion in modeling count data with excess zero responses (Zero-Inflated). FIBONACCI: Journal of Mathematics and Mathematics Education, 5(1), 71–80.https://doi.org/10.24853/fbc.5.1.71-80

Guntur, R., & Da Rato, MR (2024). Generalized Poisson regression modeling on the number of infant deaths in East Nusa Tenggara Province in 2022. J Statistika: Scientific Journal of Statistics Theory and Applications, 17(2), 779–788.https://doi.org/10.36456/jstat.vol17.no2.a9318

Hossain, S., Islam, MM, Hasan, MA, Chowdhury, PB, Easty, IA, Tusar, MK, Rashid, MB, & Bashar, K. (2023). Association of climate factors with dengue incidence in Bangladesh, Dhaka City: A calculated regression approach. Heliyon, 9(5), Article e16053.https://doi.org/10.1016/j.heliyon.2023.e16053

Ministry of Health of the Republic of Indonesia. (2016). Indonesian health profile. Ministry of Health of the Republic of Indonesia.

Suleman, RA, & Indriyani, AF (2022). Poisson Inverse Gaussian (PIG) regression on Dengue Hemorrhagic Fever (DHF) cases in Gorontalo Regency in 2022. Alpha Journal of Statistics, 1(1), 1–5.

Suryani, I., Yasin, H., & Kartikasari, P. (2021). Modeling the number of dengue hemorrhagic fever (DHF) cases in Central Java using Geographically Weighted Negative Binomial Regression (GWNBR). Gaussian Journal, 10(1), 136–148.https://doi.org/10.14710/j.gauss.v10i1.29400

Sugiyono. (2013). *Metode Penelitian Kuantitatif dan R&D*.

Sudaryono. (2022). Educational research methodology. Student Library.