西 南 交 通 大 学 学 报

第 56 卷 第 6 期
2021 年 12 月

JOURNAL OF SOUTHWEST JIAOTONG UNIVERSITY

Vol. 56 No. 6
Dec. 2021

ISSN: 0258-2724 DOI：10.35741/issn.0258-2724.56.6.88

Research article

Mathematics

COUNT DATA MODELING USING GAMLSS APPROACH AND ITS APPLICATION IN DENGUE HEMORRHAGIC FEVER CASES IN EAST KALIMANTAN PROVINCE, INDONESIA

使用 游戏机方法的計數數據建模及其在印度尼西亞東加里曼丹省登革熱出血熱病例中的應用

M. Fathurahman *, Ika Purnamasari, Surya Prangga
Department of Mathematics, Faculty of Mathematics and Natural Sciences, Mulawarman University
Samarinda, East Kalimantan, Indonesia, fathur@fmipa.unmul.ac.id

Received: September 18, 2021 ▪ Review: October 6, 2021
▪ Accepted: November 8, 2021 ▪ Published: December 24, 2021

Abstract

Generalized Additive Models for Location, Scale, and Shape (GAMLSS) is a robust approach used to model various types and characteristics of data. Therefore, this research aims to model the count data using the GAMLSS approach through Poisson Regression (PR), Poisson Inverse Gaussian Regression (PIGR), and Negative Binomial Regression (NBR). PIGR and NBR are the best models compared to PR based on their application to modeling the number of dengue hemorrhagic fever (DHF) cases in East Kalimantan Province, Indonesia, in 2019. Furthermore, both models produced varying results on the factors with a significant effect on DHF. Only one factor of the PIGR model, namely altitude, significantly affected these cases. Meanwhile, the NBR model produced three factors that affected the number of dengue cases: altitude, population density, and health workers.

Keywords: Count Data; Generalized Additive Models for Location, Scale, And Shape; Poisson Regression; Poisson Inverse Gaussian Regression; Negative Binomial Regression; Dengue Hemorrhagic Fever

摘要 位置、比例和形狀的廣義加法模型是一種強大的方法，用於對數據的各種類型和特徵進行建模。因此，本研究旨在通過泊松回歸、泊松逆高斯回歸和負二項式回歸 使用方法對計數數據進行建模。 與相比，和是最好的模型，基於它們在 2019 年印度尼西亞東加里曼丹省登革出血熱 病例數建模中的應用。此外，兩種模型在影響顯著的因素上產生了不同的結果 在上。模型中只有一個因素，即海拔高度，對這些情況有顯著影響。 同時，模型產生了影響登革熱病例數的三個因

素，如海拔、人口密度和衛生工作者。

关键词: 计数数据； 位置、比例和形状的广义加法模型； 泊松回归； 泊松逆高斯回归； 负二项式回归； 登革出血热

I. Introduction

Count data is statistical data often used in various research fields, such as social, economic, environmental, and health. The approach commonly used in its modeling process is Generalized Linear Models (GLM) [1] or Generalized Additive Models (GAM) [2]. However, there are some weaknesses associated with using these two approaches. Firstly, they can only be used to model the relationship between explanatory variables and the location parameters of the response variable distribution. Secondly, they cannot model the relationship between explanatory variables with location parameters and the shape of the response distribution. These weaknesses can be overcome using an approach called Generalized Additive Models for Location, Scale, and Shape (GAMLSS) [3].

GAMLSS is a development approach comprising GLM and GAM models that provide location parameters that describe only a limited aspect of the response variable distribution. This approach accommodates other parameters of the response variable distribution related to the explanatory variable, such as scale and shape in linear, nonlinear, parametric, semiparametric, nonparametric, and random effects functions [3, 4]. The forms of semiparametric and nonparametric were described [5, 6]. Meanwhile, the response variable in GAMLSS follows a distribution that belongs to the exponential family with the addition of discrete, continuous, skewed, and kurtosis models. It can also be used to model count data of the response variable includes underdispersion, equidispersion, and overdispersion [7].

GAMLSS was used to model the number of Dengue Hemorrhagic Fever (DHF) cases in East Kalimantan Province, Indonesia, in 2019, using the count data type. DHF is a severe infectious disease caused by the dengue virus, contaminated through the bites of the Aedes aegypti and Aedes albopictus mosquitoes. This disease disrupts the capillaries and blood-clotting systems, leading to bleeding and death while not treated properly [8]. DHF is commonly found in tropical countries, such as Indonesia, and is still a public health problem, specifically in East Kalimantan Province.

According to a publication by the Ministry of Health of the Republic of Indonesia [9], the number of positive cases and deaths in East Kalimantan Province in 2019 was 6,723 and 44, respectively. This is in addition to the Incidence Rate (IR) of 100,000 per population of 180.66 and the Case Fatality Rate (CFR) of 0.65%. Of the 34 provinces in Indonesia, East Kalimantan was ranked second for the highest IR value after North Kalimantan. IR value of DHF in this province was also very high and exceeded the overall value of 51.48 in Indonesia. This shows that the number of DHF cases in this province is very high.

This research aims to model the count data using the GAMLSS approach through PR, PIGR, and NBR to determine the best model to analyze the number of DHF cases in East Kalimantan Province, Indonesia, in 2019. It also aims to determine the factors that significantly affect the number of DHF cases. Furthermore, this research is limited to PR, PIGR, and NBR models, modeled with the GAMLSS approach and Rigby and Stasinopoulos (RS) algorithm [3, 10, 11].

II. Literature Review

A. GAMLSS

GAMLSS assumes that the response variable is Y_i for $i = 1,2,\dots,n$ with a probability distribution function $P(Y_i = y_i|\boldsymbol{\theta}^i)$ where $\boldsymbol{\theta}^i = [\theta_{i1} \quad \theta_{i2} \quad \cdots \quad \theta_{ip}]^T$. $\boldsymbol{\theta}^i$ is a vector of four distribution parameters, namely $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$, $\boldsymbol{\upsilon}$, and $\boldsymbol{\tau}$ which are functions of explanatory variables. The parameters $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are referred to as location and scale, while $\boldsymbol{\upsilon}$ and $\boldsymbol{\tau}$ are known as skewness and kurtosis and included in the shape parameters [3].

Supposing $\boldsymbol{y}^T = [y_1 \quad y_2 \quad \cdots \quad y_n]$ denotes a vector of response variables with a size of $(n \times 1)$ and $g_k(\cdot), k = 1,2,3,4$ is a connecting function between distribution parameters and explanatory variables. GAMLSS model can be written as follows [3]:

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = \boldsymbol{X}_k\boldsymbol{\beta}_k + \sum_{j=1}^{J_k} \boldsymbol{Z}_{jk}\gamma_{jk} \qquad (1)$$

Supposing $\boldsymbol{Z}_{jk} = \boldsymbol{I}_n$, where $\boldsymbol{I}_n$ is an identity matrix with a size of $n \times n$ and $\gamma_{jk} = h_{jk} = h_{jk}(\boldsymbol{x}_{jk})$ for all combinations of j and k in Equation (1), then GAMLSS model can be rewritten as follows [3]:

$$
\begin{aligned}
g_k(\boldsymbol{\theta}_k) &= \boldsymbol{\eta}_k = \boldsymbol{X}_k\boldsymbol{\beta}_k + \sum_{j=1}^{J_k} h_{jk}(\boldsymbol{x}_{jk}) \\
g_1(\boldsymbol{\mu}) &= \boldsymbol{\eta}_1 = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \sum_{j=1}^{J_1} h_{j1}(\boldsymbol{x}_{j1}) \\
g_2(\boldsymbol{\sigma}) &= \boldsymbol{\eta}_2 = \boldsymbol{X}_2\boldsymbol{\beta}_2 + \sum_{j=1}^{J_2} h_{j2}(\boldsymbol{x}_{j2}) \qquad (2) \\
g_3(\boldsymbol{v}) &= \boldsymbol{\eta}_3 = \boldsymbol{X}_3\boldsymbol{\beta}_3 + \sum_{j=1}^{J_3} h_{j3}(\boldsymbol{x}_{j3}) \\
g_4(\boldsymbol{\tau}) &= \boldsymbol{\eta}_4 = \boldsymbol{X}_4\boldsymbol{\beta}_4 + \sum_{j=1}^{J_4} h_{j4}(\boldsymbol{x}_{j4})
\end{aligned}
$$

where:

$\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{v}, \boldsymbol{\tau}, \boldsymbol{\eta}_k$ = vector with length n;

$\boldsymbol{\beta}_k^T = [\beta_{1k} \quad \beta_{2k} \quad \cdots \quad \beta_{J'_k k}]$ is the parameter vector;

$\boldsymbol{X}_k$ = explanatory variable matrix of size $n \times J'_k$;

h_{jk} = nonparametric smoothing function of the explanatory variables $\boldsymbol{x}_{jk}$, where $j = 1,2,\ldots,j_k$ is also a vector of length n and $k = 1,2,3,4$. The function h_{jk} is the unknown function of the explanatory variables X_{jk}, and $h_{jk} = h_{jk}(\boldsymbol{x}_{jk})$ is a vector that evaluates the function h_{jk} on x_{jk}.

Inference to the GAMLSS model includes parameter estimation and hypothesis testing. Parameter estimators can be obtained using the penalized likelihood method and numerical approach. These include the RS algorithm, Cole and Green (CG) algorithm, and a mixture of RS and CG algorithms (RS-CG algorithm) [10,11]. However, this research only uses the RS algorithm, while the parameter hypothesis testing was conducted with the likelihood ratio test and Wald test procedures [12].

B. Distribution of Count Data on GAMLSS

Some distributions for modeling count data using the GAMLSS approach are as follows:

1) Poisson Distribution

This is a distribution of discrete random variables that express the number of successes of an experiment. According to Cameron & Trivedi [13], it has the following characteristics:

- An event with a small probability occurs in a population with many members.
- Depends on a certain time interval.
- Events are included in the stochastic process.
- The recurrence of events follows the binomial distribution.

Let Y be a discrete random variable with Poisson distribution consisting of parameter μ. Then the probability mass function is obtained as follows [13, 14]:

$$P(Y = y|\mu) = \frac{e^{-\mu}\mu^y}{y!}, y = 0,1,2,\ldots \qquad (3)$$

where μ is a location parameter representing the probability of many successful events from Y. Poisson distribution has the same mean and variance, namely $E(Y) = \mu$ and $Var(Y) = \mu$. Meanwhile, the skewness is $S = E(Y-\mu)^3/\mu^3 = 1/\sqrt{\mu}$.

2) Poisson Inverse Gaussian Distribution

This is a mixed Poisson distribution with its shape depending on the random effect (v). Suppose $q(v)$ denotes the integral υ obtains the probability density function of v and the marginal distribution for Y:

$$P(Y = y|\mu) = \int f(y|\mu, v)\, q(v) dv, \qquad (4)$$

where v is assumed to have an inverse Gaussian distribution and a probability density function written as follows:

$$q(v) = (2\pi\tau v^3)^{-\frac{1}{2}} e^{-(v-1)^2/2\tau v}, \quad v > 0 \qquad (5)$$

where $E(V) = 1$ and $\tau = Var(V)$. Based on Equations (4) and (5), PIG distribution is formed with the probability mass function as follows [15-17]:

$$P(Y = y|\mu,\tau) = \left(\frac{2z}{\pi}\right)^{1/2} \frac{\mu^y e^{1/\tau} K_s(z)}{(z\tau)^y y!}, \qquad (6)$$

where $s = y - 1/2$, $z = \sqrt{1/\tau^2 + 2\mu/\tau}$, and $K_s(z) = K_{y-1/2}(1/\tau\sqrt{2\mu\tau + 1})$ is a modified Bassel function of the third kind.

3) Negative Binomial Distribution

Negative binomial distribution has two parameters, namely μ andσ, which denotes the parameter location and dispersion. The probability mass function of the negative binomial distribution is formulated as follows:

$$P(Y = y|\mu,\sigma) = \frac{\Gamma((y+1)/\sigma)\alpha^y}{\Gamma(y+1)\Gamma(1/\sigma)} \left(\frac{(\mu\sigma)^y}{\mu\sigma + 1}\right)^{y+\frac{1}{\sigma}}, \qquad (7)$$

$y = 0,1,2,\ldots,\infty$

where $E(Y) = \mu$ is the mean, $Var(Y) = \mu(1+\alpha)$ is the variance, and α is the dispersion parameter [18,19].

C. Best Model Selection

The criteria that can be used to obtain the best regression model based on the GAMLSS approach are Global Deviance (GDEV) and Akaike Information Criterion (AIC), which are defined as follows:

$$\begin{aligned} \text{GDEV} &= -2\ell(\hat{\boldsymbol{\theta}}), \\ \text{AIC} &= -2\ell(\hat{\boldsymbol{\theta}}) + 2K, \end{aligned} \tag{8}$$

where $\ell(\hat{\boldsymbol{\theta}})$ denotes the maximum log-likelihood of the model and K is the number of parameters in the model. The best models have the smallest GDEV and AIC values [11].

III. RESEARCH METHODS

This research uses secondary data from the Central Statistics Agency and the Provincial Health Office of East Kalimantan, Indonesia [20,21]. It uses one response variable (Y) and four explanatory variables (X_1, X_2, X_3, X_4), as shown in Table 1.

Table 1.
Research variables

Symbol	Variable	Variable Type
X_1	An area	Continuous
X_2	Area altitude	Continuous
X_3	Population density	Continuous
X_4	Number of health workers	Discrete
Y	Number of DHF cases	Discrete

The steps of data analysis in this research are as follows:

1. Performing descriptive statistical analysis of research data.
2. Performing multicollinearity detection of explanatory variables.
3. Modeling the number of DHF cases using PR, PIGR, and NBR models based on the GAMLSS approach.
4. Getting the best model needed for the number of DHF cases.
5. Getting the factors that influence DHF cases.
6. Concluding.

IV. RESULTS AND DISCUSSION

The discussion starts with a descriptive analysis of research data, as shown in Table 2.

Table 2.
Description of research data

Variable	Mean	SD a	Max b	Min c
X_1	12,734.7	11,494.05	31,051.7	163.1
X_2	54.1	65.04	174.63	5.98
X_3	380	579	1,298	1
X_4	1,381	903	2,960	287
Y	672	621	1,838	66

[a] Standard deviation, [b] maximum, [c] minimum

Table 2 shows that the average number of DHF cases at East Kalimantan Province in 2019 was 672. The highest and lowest numbers, 1,838 and 6 cases, were found in Balikpapan City and Mahakam Ulu Regency, respectively. The data description on the number of DHF cases is shown in Figure 1. The average area of regencies/cities in East Kalimantan Province is 12,374.7 km^2, while the largest and smallest areas are found in Kutai Timur Regency (31,051.7 km^2) and Bontang City (163.1 km^2). The average height of 54.1 meters and height altitude of 174.63 masl are found in the regency/city in East Kalimantan Province and Mahakam Ulu Regency, respectively.

Meanwhile, the regency/city with the lowest altitude in Kutai Timur Regency is at 31,051.7 km^2 with an average population density of 380 people/km^2. The highest population density is in Balikpapan City, with 1,298 people/km^2, and the lowest is in Mahakam Ulu Regency, as much as one person/km^2. Furthermore, the average number of health workers in this province in 2019 was 1,381 people, with the highest and lowest found in Samarinda City (2,960 people) and Mahakam Ulu Regency (287 people), respectively.

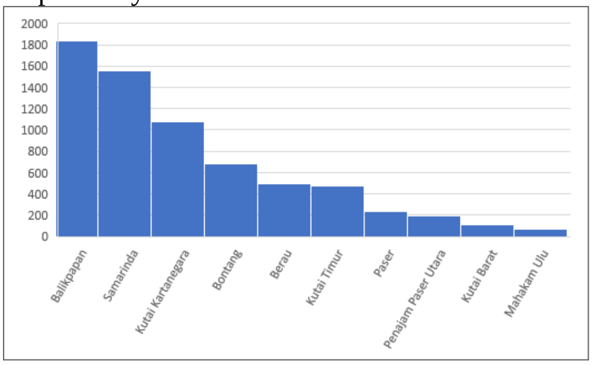

Figure 1. The description of the number of DHF cases in East Kalimantan Province, Indonesia, in 2019

This is followed by testing the prerequisites for data analysis of PR, PIGR, and NBR models, namely detecting collinearity between explanatory variables. Multicollinearity detection uses the Variance Inflation Factor (VIF) value with an explanatory variable comprising a VIF value of more than 10 [22], as shown in Table 3.

Table 3.
VIF value of explanatory variables

Explanatory Variables	VIF Value
X_1	2.89
X_2	1.35
X_3	5.15
X_4	2.99

Table 3 shows that all explanatory variables have a VIF value of less than 10. These results indicate that there is no multicollinearity in the explanatory variables. Therefore, it is important to use PR, PIGR, and NBR models.

The results of modeling the number of DHF cases in East Kalimantan Province, Indonesia, in 2019, using PR, PIGR, and NBR models with the GAMLSS approach are shown in Table 4.

Table 4.
Estimation and testing of PR, PIGR, and NBR model parameters

Model	Par a	Est b	Se c	Z	p-value
PR	β_0	4.93	6.62×10-2	74.50	8.26×10-9 d
	β_1	3.24×10-5	2.67×10-6	12,11	6.77×10-5 d
	β_2	-5.26×10-3	4.24×10-4	-12.41	6.01 ×10-5 d
	β_3	9.75×10-4	6.66×10-5	14.64	2.69×10-5 d
	β_4	4.93×10-4	2.52×10-5	19.55	6.45×10-6 d
PIGR	β_0	5.01	0.27	18.28	5.27×10-5 d
	β_1	2.68×10-5	1.78×10-5	1.51	0.2166
	β_2	-6.72×10-3	1.97×10-3	-3.41	0.0271 d
	β_3	9.69×10-4	5.86×10-4	1.65	0.1735
	β_4	5.19×10-4	3.13×10-4	1.66	0.1724
NBR	β_0	4.99	2.33×10-1	21.45	2.79×10-5 d
	β_1	2.76×10-5	1.51×10-5	1.82	0.1423
	β_2	-6.79×10-3	1.68×10-3	-4.03	0.0157 d
	β_3	9.61×10-4	2.70×10-4	3.56	0.0237 d
	β_4	2.24×10-4	3.49×10-5	15.00	0.0001 d

[a] Parameter, [b] estimate, [c] standard error, [d] significant at the level of significance, α = 0.05

Table 4 shows that all PR model parameters are significant at a significance level of 0.05. Furthermore, for the PIGR model, only two parameters are significant, namely β_0 and β_2. Meanwhile, the significant parameters in the NBR model are four, namely β_0, β_2, β_3, and β_4.

Furthermore, this research selects the best model needed to determine the number of DHF cases in East Kalimantan Province in 2019. The results obtained are shown in Table 5.

Table 5.
Selection of the best model with GDEV and AIC values

Model	GDEV	AIC
PR	416.79	426.79
PIGR	122.04 [a]	134.04 [a]
NBR	122.29	134.29

[a] Best model

Based on Table 5, the model that has the smallest GDEV and AIC is PIGR. These results indicate that the best model for analyzing the number of DHF cases in East Kalimantan Province in 2019 is the PIGR model. Meanwhile, the values of GDEV and AIC for the NBR model are relatively the same or close to PIGR. Therefore, it is necessary to consider selecting the NBR model as an alternative to analyzing the number of DHF cases in East Kalimantan Province in 2019, specifically the NBR model with many significant parameters compared to PIGR.

PIGR model is obtained as follows:

$$\log(\hat{\mu}) = 5.008 + 2.68 \times 10^{-5} X_1 - 6.724 \times 10^{-3} X_2 + 9.685 \times 10^{-4} X_3 + 5.185 \times 10^{-4} X_4.$$

The significant explanatory variable in the PIGR model above is the altitude of the region (X_2), which is used to interpret regencies/cities with a tendency of reducing the number of dengue cases by 1.0067 times.

V. Conclusion

GAMLSS is a flexible approach adaptable to various characteristics of data or distributions.

This approach can accommodate other parameters from the distribution of response variables related to explanatory variables. These examples are scale and shape parameters in linear, nonlinear, parametric, nonparametric, and random effects. The response variable in GAMLSS follows a distribution that belongs to the exponential family and includes discrete and continuous distributions with highly skewed and kurtosis. Therefore, the GAMLSS approach can be used to model count data.

Based on the application of GAMLSS through PR, PIGR, and NBR models in analyzing the number of DHF cases in East Kalimantan Province, Indonesia, in 2019, PIGR was obtained as the best. However, the NBR model can be an alternative because it has GDEV and AIC values as PIGR. NBR produces more significant parameters than PIGR. The area's height is the only factor that significantly affects the number of DHF cases in East Kalimantan Province in 2019 based on the PIGR model. Meanwhile, NBR produces three factors that significantly affect the area's altitude, population density, and health workers. PIGR and NBR models produce a regional altitude factor that significantly affects the number of DHF cases in East Kalimantan Province, Indonesia, in 2019.

ACKNOWLEDGMENT

We would like to thank Mulawarman University for funding the research through IsDB Grant Research under 405/UN17.11/PL/2021.

REFERENCES

[1] NELDER, J.A., and WEDDERBURN, D.M. (1972) Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, 135 (3), pp. 370-384.
[2] HASTIE, T.J., and TIBSHIRANI, R.J. (1990) *Generalized additive models*. New York: Chapman and Hall/CRC.
[3] RIGBY, R.A., and STASINOPOULOS, D.M. (2005) Generalized additive models for location, scale, and shape (with discussion). *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54 (3), pp. 507-554.
[4] AEBERHARD, W.H., CANTONI, E., MARRA, G., and RADICE, R. (2021) Robust fitting for generalized additive models for location, scale, and shape. *Statistics and Computing*, 31 (5), pp. 1-16.
[5] LESTARI, B., and CHAMIDAH, N. (2020) Estimating regression function of multi-response semiparametric regression model using smoothing spline. *Journal of Southwest Jiaotong University*, 55 (5), pp. 1-9.
[6] SEDIONO, MARDIANTO, M.F.F., ULYAH, S.M., ARDHANI, B.A., APRILIANTI, N.A., and RAHMADINA, R.F. (2020) AMiBI: Application of flood mitigation in Indonesia based on the results of statistical analyses of causal factors using local linear estimators in the nonparametric regression model. *Journal of Southwest Jiaotong University*, 55 (6), pp. 1-8.
[7] RIGBY, R.A., STASINOPOULOS, D.M., HELLER, G., and DE BASTIANI, F. (2020) *Distributions for modeling location, scale, and shape using GAMLSS in R*. Boca Raton: Chapman and Hall/CRC.
[8] WORLD HEALTH ORGANIZATION (2011) *Comprehensive guidelines for preventing and controlling dengue and dengue hemorrhagic fever*. New Delhi: World Health Organization Regional Publication SEARO.
[9] THE MINISTRY OF HEALTH OF THE REPUBLIC OF INDONESIA. (2020) *Health profile of Indonesia 2019*. Jakarta: The Ministry of Health of the Republic of Indonesia.
[10] STASINOPOULOS, D.M., and RIGBY, R.A. (2007) Generalized additive models for location, scale, and shape (GAMLSS) in R. *Journal of Statistical Software*, 23 (7), pp. 1-46.
[11] STASINOPOULOS, D.M., RIGBY, R.A., HELLER, G., VOUDOURIS, V., and DE BASTIANI, F. (2017) *Flexible regression and smoothing: using GAMLSS in R*. Boca Raton: Chapman and Hall/CRC.
[12] PAWITAN, Y. (2001) *In all likelihood: Statistical modeling and inference using likelihood*. Oxford: Clarendon Press.
[13] CAMERON, A.C. and TRIVEDI, P.K. (2013) *Regression analysis of count data*. 2nd ed. New York: Cambridge University Press.
[14] AGRESTI, A. (2013) *Categorical data analysis*. 3rd ed. New York: John Wiley & Sons.

[15] HOLLA, M.S. (1966) On a Poisson-inverse-Gaussian distribution. *Metrika*, 11, pp. 115-121.
[16] WILMOT, G.E. (1987) The Poisson-inverse Gaussian distributions as an alternative to the negative binomial. *Scandinavian Actuarial Journal*, pp. 113-127.
[17] DEAN, C., LAWLESS, J.F., and WILMOT, G.E. (1989) A mixed Poisson-inverse-Gaussian regression model. *The Canadian Journal of Statistics*, 17 (2), pp. 171-181.
[18] LAWLESS, J.F. (1987) Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics*, 15 (3), pp. 209-225.
[19] HILBE, J.M. (2011) *Negative binomial regression*. 2nd ed. New York: Cambridge University Press.
[20] THE CENTRAL STATISTICS AGENCY (2020) *East Kalimantan Province in figures 2020.* Samarinda: The Central Statistics Agency of the East Kalimantan Province, Indonesia.
[21] THE PROVINCIAL HEALTH OFFICE OF EAST KALIMANTAN, INDONESIA (2020) *Health profile of East Kalimantan Province 2019.* Samarinda: The Provincial Health Office of East Kalimantan, Indonesia.
[22] KUTNER, M., NACHTSHEIM, C.J., and NETER, J. (2004) *Applied linear regression models*. New York: McGraw-Hill.

参考文:

[1] NELDER, J.A., 和 WEDDERBURN, D.M. (1972) 廣義線性模型。皇家統計學會雜誌：一个系列（綜合），135 (3)，第 370-384 頁。
[2] HASTIE, T.J., 和 TIBSHIRANI, R.J. (1990) 廣義加法模型。紐約：査普曼和霍爾/冷库。
[3] RIGBY, R.A., 和 STASINOPOULOS, D.M. (2005) 位置、比例和形狀的廣義加法模型（帶討論）。皇家統計學會雜誌：C 系列（應用統計），54 (3)，第 507-554 頁。
[4] AEBERHARD, W.H., CANTONI, E., MARRA, G., 和 RADICE, R. (2021) 位置、比例和形狀的廣義加法模型的穩健擬合。統計與計算，31 (5)，第 1-16 頁。
[5] LESTARI, B., 和 CHAMIDAH, N. (2020) 使用平滑樣條估計多響應半參數回歸模型的回歸函數。西南交通大學學報，55（5），1-9頁。
[6] SEDIONO, MARDIANTO, M.F.F., ULYAH, S.M., ARDHANI, B.A., APRILIANTI, N.A., 和 RAHMADINA, R.F.（2020）AMiBI：基於非參數回歸模型中使用局部線性估計的因果因素統計分析結果在印度尼西亞的防洪應用。西南交通大學學報，55（6），1-8頁。
[7] RIGBY, R.A.、STASINOPOULOS, D.M.、HELLER, G., 和 DE BASTIANI, F. (2020) 在·博卡拉頓中使用 游戏机建模位置、比例和形狀的分佈：査普曼和 大廳/冷库。
[8] 世界衛生組織（2011）登革熱和登革出血熱防控綜合指南。新德里：世界衛生組織區域出版物 西羅。
[9] 印度尼西亞共和國衛生部。(2020) 2019 年印度尼西亞健康概況。雅加達：印度尼西亞共和國衛生部。
[10] STASINOPOULOS, D.M., 和 RIGBY, R.A. (2007) R 中位置、比例和形狀的廣義加法模型。統計軟件雜誌，23 (7)，第 1-46 頁。
[11] STASINOPOULOS, D.M., RIGBY, R.A., HELLER, G., VOUDOURIS, V., 和 DE BASTIANI, F. (2017) 靈活回歸和平滑：在 R 中使用。博卡拉頓：査普曼和霍爾/冷库。
[12] PAWITAN, Y. (2001) 很有可能：使用可能性的統計建模和推理。牛津：克拉倫登出版社。
[13] CAMERON, A.C., 和 TRIVEDI, P.K. (2013) 計數數據的回歸分析。第 2 版。紐約：劍橋大學出版社。
[14] AGRESTI, A. (2013) 分類數據分析。第 3 版。紐約：約翰威利父子公司。
[15] HOLLA，M.S. (1966) 關於泊松逆高斯分佈。指標，11，第 115-121 頁。
[16] WILMOT，G.E. (1987) 泊松逆高斯分佈作為負二項式的替代。斯堪的納維亞精算雜誌，第 113-127 頁。

[17] DEAN, C., LAWLESS, J.F., 和 WILMOT, G.E. (1989) 混合泊松逆高斯回歸模型。加拿大統計雜誌，17 (2)，第 171-181 頁。
[18] LAWLESS, J.F. (1987) 負二項式和混合泊松回歸。加拿大統計雜誌，15 (3)，第 209-225 頁。
[19] HILBE, J.M. (2011) 負二項式回歸。第 2 版。紐約：劍橋大學出版社。
[20] 中央統計局（2020 年）東加里曼丹省，2020 年數據。三馬林達：印度尼西亞東加里曼丹省中央統計局。
[21] 印度尼西亞東加里曼丹省衛生辦公室（2020）2019 年東加里曼丹省衛生概況。三馬林達：印度尼西亞東加里曼丹省衛生辦公室。
[22] KUTNER, M.、NACHTSHEIM, C.J., 和 NETER, J. (2004) 應用線性回歸模型。紐約：麥格勞-希爾。