Indonesian Journal of Applied Mathematics and Statistics (IdJAMS)
ISSN 3064-4062
Volume 3 Issue 1 June, 2026, pp. 21-26
https://idjams.com/index.php/jms/index

Modeling Diabetes Mellitus Prevalence in West Java Using Negative Binomial Regression

Elaiya Najla Oasis[1,*], Nasywa Firdausi Chandra[2], Taskya Prianggita Alfinayah[3], Abdul Azis Abdillah[4], Restu Arisanti[5]

[1, 2, 3] *Statistics Study Program, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Jl. Soekarno Km 21 Jatinangor, Kab. Sumedang, West Java 45363, Indonesia*

[4] *The School of Engineering, University of Birmingham, Birmingham B15 2TT, United Kingdom*

[5] *Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Jl. Soekarno Km 21 Jatinangor, Kab. Sumedang, West Java 45363, Indonesia*

[*]*Corresponding author: elaiya22001@mail.unpad.ac.id*

Copyright: ©2026 by Indonesian Journal of Applied Mathematics and Statistics. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution-ShareAlike 4.0 International License-(CC-BY-SA) (https://creativecommons.org/licenses/by-sa/4.0/

Article Info

Article history:
Received: November 20, 2025
Revised: December 20, 2025
Accepted: December 21, 2025
Available online: June 1, 2026

DOI:
https://doi.org/10.71385/idjams.v3i1.31

Abstract

Diabetes Mellitus is a disease with a high number of sufferers in Indonesia. According to the West Java Provincial Health Office, the number of people with Diabetes Mellitus in West Java Province in 2022 reached 644,704 cases. The causative factor of Diabetes Mellitus is obesity, one of the triggers is food intake. Foods consisting of carbohydrates, protein, saturated fat, sugar, and fat in general can increase blood glucose levels and increase the risk of diabetes. This study aims to determine the role of food in the prevalence of Diabetes Mellitus in West Java. The method used was negative binomial regression analysis. The results obtained show that the level of consumption of fruits and vegetables, oils and fats, and nuts has a significant influence on the prevalence of Diabetes Mellitus in West Java. Public awareness to implement good eating behavior is very important to reduce the prevalence of Diabetes Mellitus in West Java.

Keywords: Diabetes Mellitus; Food Consumption; Negative Binomial; Poisson

I. Introduction

Diabetes mellitus (DM) is a condition caused by uncontrolled elevated blood glucose levels due to insulin deficiency, insulin resistance, or both. Insulin is a hormone produced by pancreatic β-cells that regulates blood glucose by controlling glucose use and storage. Insulin deficiency mainly occurs due to damage to pancreatic β-cells, which are responsible for insulin production. Additionally, DM can also be caused by insulin resistance. Insulin resistance is a reduced ability of insulin to stimulate glucose utilization or a decrease in the response of target cells, such as muscles, tissues, and the liver, to physiological insulin levels [1]. DM is divided into type 1 diabetes, type 2 diabetes, maturity-onset diabetes of the young (MODY), gestational diabetes, neonatal diabetes, and steroid-induced diabetes [2]. According to the Indonesian Ministry of Health in 2014, DM is known as a dangerous silent killer because it is often not realized by sufferers and begins to be realized when it has complications. DM can attack from the skin to the heart and cause complications [3]. DM also impacts the lives and well-being of individuals, families, and communities [1].

In 2022, World Health Organization (WHO) reported that 422 million people worldwide were living with DM. From all over the world, China has the highest number of sufferers, followed by India, Pakistan, the United States, and Indonesia [4]. Based on the International Diabetes Federation (IDF) report, in 2022, there were around 41,817 people suffering from type 1 diabetes in Indonesia, making it

Indonesian Journal of Applied Mathematics and Statistics (IdJAMS)
ISSN 3064-4062
Volume 3 Issue 1 June, 2026, pp. 21-26
https://idjams.com/index.php/jms/index

the highest in ASEAN [5] [6]. At the regional level, West Java Provincial Health Office reported 644,704 DM cases in 2022, with the highest number reported in Bogor Regency. These figures indicate that DM remains a significant health burden in West Java [7].

Risk factors for DM include genetic, environmental, age, race, lifestyle, obesity, and dietary factors. Diet is a modifiable factor and plays an important role in diabetes prevention. Consumption of fruits, vegetables, and nuts is associated with a reduced risk of diabetes, although it also affects total caloric intake, whereas excessive intake of oils and fats, particularly saturated fats, is associated with an increased risk of diabetes mellitus [8] [1]. Consuming foods with a low glycemic index and increasing fiber intake can help reduce diabetes risk [8]. Both water-soluble fiber and water-insoluble fiber are found in foodstuffs such as fruits, vegetables, and nuts. Both types of fiber work together to reduce postprandial blood sugar and triglyceride levels and decrease the postprandial insulin-glucose ratio via Peroxisome Proliferator Activated Receptor (PPAR) [9]. A balanced diet also contributes to glycemic control and reductions in HbA1c [1].

Given the importance of diet in diabetes prevention, this study focuses on food-related factors, namely consumption of fruits and vegetables, nuts, and oils and fats. The response variable in this study is the number of people with diabetes mellitus in West Java Province in 2022, which represents count data and follows a Poisson distribution with potential overdispersion. Therefore, Poisson regression analysis is used to examine the relationship between dietary factors and diabetes incidence. The findings of this study are expected to support evidence-based dietary interventions and contribute to diabetes prevention efforts in West Java Province.

II. Research Methods

A. Data

The data used is secondary data obtained indirectly from the object of research. Secondary data sources can come from the same reference or from a site [10]. Secondary data in this study are data on the number of people with diabetes mellitus in West Java Province in 2022, the level of consumption of fruits and vegetables in West Java Province in 2022, the level of consumption of oils and fats in West Java Province in 2022, and the level of consumption of legumes in West Java Province in 2022. The variables to be studied are Number of Diabetes Mellitus Patients, Fruit and Vegetable Consumption, Oil and Fat Consumption, and Nuts Consumption with the Number of DM Patients as the response variable and Fruit and Vegetable Consumption, Oil and Fat Consumption, and Nuts Consumption as predictor variables. The research data variables are as follows:

Table 1. Research Data Variables

Data	Unit of Measurement	Variable
Number of Diabetes Mellitus Patients	Count	Y
Fruit and Vegetable Consumption	kg/capita	X_1
Oil and Fat Consumption	kg/capita	X_2
Nuts Consumption	kg/capita	X_3

B. Poisson Regression Analysis

Poisson regression analysis is a regression model that allows it to be used on data where the response variable is not normally distributed and discrete. This model is an implementation of the Generalized Linear Model (GLM), which shows the relationship between the predictor variable X and the response variable Y of discrete data with Poisson distribution [11]. According to Montgomery, the model of Poisson regression can be written as follows:

$$y_i = E(y_i) + \varepsilon_i, i = 0,1,\ldots,n, \;\; E(y_i) = \mu_i \tag{1}$$

where yi is the response variable for the i-th observation, E(yi) is the mean of the response variable for the i-th observation, and i are the errors. The link function that can be used:

Indonesian Journal of Applied Mathematics and Statistics (IdJAMS)
ISSN 3064-4062
Volume 3 Issue 1 June, 2026, pp. 21-26
https://idjams.com/index.php/jms/index

$$g(\mu_i) = \eta_i = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k \quad (2)$$

So the form of the Poisson regression model can be written as follows [12]:

$$y_i = exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k) + \varepsilon_i \quad (3)$$

To determine whether the Poisson distribution is appropriate for the count data, equidispersion was assessed by calculating the dispersion index, denoted by φ, which was calculated as the ratio of the model deviance to its degrees of freedom [13]. Excessive dispersion can lead to biased parameter estimates and invalid conclusions. The Poisson regression model must be equidispersion, where the mean and variance of the response variable are equal, while overdispersion means that the variance value is greater than the mean, making the Poisson regression model is unsuitable for the data [14] [15].

C. Negative Binomial Regression Analysis

Negative Binomial regression is one of the applied regression models of the Generalized Linear Model (GLM) [14]. The Negative Binomial regression model is formed from the Negative Binomial distribution. If overdispersion occurs, Negative Binomial Regression is one of the models that can be used to handle this because Negative Binomial Regression does not emphasize the assumption of equidispersion contained in Poisson Regression. The general form of Negative Binomial Regression is [16] [17]:

$$\mu_i = exp(X\beta), i = 1,2,\ldots,n \quad (4)$$

where μ_i are the response variable that Negative Binomial distributed and vector-shaped with size $(n \times 1)$, X is a predictor variable in the form of a matrix with size $(n \times (p+1))$, and is a parameter that is a vector with size $((p+1) \times 1)$.

III. Result and Discussion

A. Descriptive Analysis

Table 2. Descriptive Statistics of Variables

Variables	Mean	Maximum	Minimum
Y	23,877.92593	63,978	2,377
X_1	84.56296296	116.3	664
X_2	9.637037037	11.5	7.8
X_3	12.25185185	17.5	9.4

Based on Table 2, it can be seen that the average number of people with Diabetes Mellitus (Y) in West Java Province in 2022 is around 23,878 people. The highest number of people with Diabetes Mellitus is in Bogor Regency, which reaches 63,978 people, while the lowest number of people with Diabetes Mellitus is in Banjar City with 2,377 people. The average level of fruit and vegetable consumption (X_1) in West Java Province in 2022 was 84.563 kg. The highest level of fruit and vegetable consumption is in Indramayu Regency with 116.3 kg, while the lowest level of fruit and vegetable consumption is in Bandung Regency with 66.4 kg. The average consumption level of oil and fat (X_2) in West Java Province in 2022 was 9.6 kg. The highest level of oil and fat consumption is in Bekasi City with 11.5 kg, while the lowest level of oil and fat consumption is in Tasikmalaya City with 7.8 kg. The average consumption level of nuts (X_3) in West Java Province in 2022 was 12.25 kg. The highest level of nuts consumption is in Indramayu Regency with 17.5 kg, while the lowest level of nuts consumption is in Cianjur Regency with 9.4 kg.

B. Poisson Regression Analysis

Before analyzing using Poisson Regression analysis, it is necessary to test the multicollinearity assumption to determine whether there is a correlation between variables in the data. Multicollinearity

Indonesian Journal of Applied Mathematics and Statistics (IdJAMS)
ISSN 3064-4062
Volume 3 Issue 1 June, 2026, pp. 21-26
https://idjams.com/index.php/jms/index

was assessed using the Variance Inflation Factor (VIF), and since all VIF values were less than 10, it was concluded that there was no evidence of high correlation between the independent variables in the model.

Table 3. VIF Values

Variables	Values
X_1	1.368064
X_2	1.045512
X_3	1.414045

Because no multicollinearity was detected among the explanatory variables, further analysis was conducted using Poisson regression.

Table 4. Poisson Regression Parameter Estimation Results

Variables	Estimate	Std. Error	Z Value	P-Value
(Intercept)	7.4083601	0.0158208	468.3	$< 2 \times 10^{-16}$
X_1	-0.0160604	0.0001374	-116.9	$< 2 \times 10^{-16}$
X_2	0.2874730	0.0013999	205.4	$< 2 \times 10^{-16}$
X_3	0.0979074	0.0007722	126.8	$< 2 \times 10^{-16}$

Based on the above results, the Poisson Regression model is formed as follows:

$$\hat{\lambda} = exp(7.4083601 - 0.0160604x_1 + 0.2874730x_2 + 0.0979074x_4)$$

To ensure the suitability of the Poisson Regression model, it is necessary to conduct an Overdispersion Test using the dispersion index (φ), which is defined as the ratio of the model's residual deviance to its degrees of freedom. Based on the calculation, the dispersion index test results show a value of 7220.28, which indicates the presence of overdispersion in the data since the value is greater than 1. Although the Poisson regression model was estimated, substantial overdispersion was detected, meaning that the equidispersion assumption of the Poisson model was violated.

Because the equidispersion assumption was not satisfied, the Poisson regression model was considered unsuitable for these data. To address the overdispersion problem and obtain reliable parameter estimates, Negative Binomial regression analysis was carried out as an alternative model capable of accommodating the observed overdispersion.

C. Negative Binomial Regression Analysis

The negative binomial regression model is a flexible and ideal alternative because it does not require equal means and variances. When overdispersion is present, negative binomial regression can provide a better fit for the data and produce more accurate results.

Table 5. Parameter Estimation Results of Negative Binomial Regression

Variables	Estimate	Std. Error	Z Value	P-Value
(Intercept)	7.41639	1.25950	5.888	3.9×10^{-96}
X_1	-0.02885	0.01004	-2.873	0.00407
X_2	0.32076	0.10859	2.954	0.00314
X_3	0.15775	0.05913	2.668	0.00763

Based on the calculation results, it is found that the $p - value < \alpha$ for all variables. This indicates that all variables have a significant influence, and the Negative Binomial Regression model is obtained as follows:

$$\mu_i = exp(7.41639 - 0.02885\ X_1 + 0.32076\ X_2 + \ 0.15775\ X_3)$$

D. Determination of the Best Model

Determination of the best model is done by comparing the AIC (Akaike Information Criterion) values of the two regression models. The calculation results are as follows:

Table 6. Comparison of AIC Values

Regression Model	AIC
Poisson	166,390
Negative Binomial	586.62

The markedly lower AIC value of the negative binomial model indicates that the Poisson model is severely misspecified due to extreme overdispersion. By incorporating an additional dispersion parameter, the negative binomial model offers a more suitable specification for the count data. Therefore, the Negative Binomial regression model with an AIC value of 586.62 is considered suitable for modeling the role of food in the prevalence of Diabetes Mellitus in West Java.

IV. Conclusion

This study employed a negative binomial regression model to examine the association between food consumption and the prevalence of Diabetes Mellitus in West Java with the following modeling results:

$$\mu_i = exp(7.41639 - 0.02885\ X_1 + 0.32076\ X_2 + \ 0.15775\ X_3)$$

The negative coefficient on X_1 indicates that the higher the level of fruit and vegetable consumption, the lower the likelihood of diabetes mellitus, while the positive coefficients on X_2 and X_3 indicate that the higher the level of oil, fat, and nuts consumption, the higher the likelihood of diabetes mellitus.

From a methodological perspective, the negative binomial regression model substantially outperformed the Poisson model, as indicated by a large reduction in the Akaike Information Criterion (AIC). This drastic difference reflects severe overdispersion in the data, which violates the equidispersion assumption of the Poisson model and leads to model misspecification. By explicitly modeling extra-Poisson variation through an additional dispersion parameter, the negative binomial model provides a more appropriate and reliable representation of the underlying count data.

However, several limitations must be acknowledged. First, this study used an ecological design with provincial-level data, which limited the ability to draw causal conclusions at the individual level. Second, data on lifestyle behaviors, particularly physical activity and exercise, were not available. This causes the analysis to be limited to only three predictors of consumption patterns, so that other important factors such as physical activity, obesity, smoking behavior, and socioeconomic status were not used in the modeling.

Despite these limitations, the findings provide relevant policy implications. The results indicate that fruit and vegetable consumption is negatively associated with diabetes prevalence, whereas oil, fat, and nuts consumption show positive associations. Regional health authorities should promote increased consumption of fruits and vegetables as part of balanced dietary guidelines, while encouraging the reduction of excessive oil and fat intake. Although nuts are high in fiber and are not a major factor in the high prevalence of diabetes, consumption of nuts should still not be excessive. In addition, strengthening regional nutrition education programs and integrating dietary interventions into existing public health initiatives may help reduce the prevalence of Diabetes Mellitus in West Java.

References

[1] D. Hardianto, "Telaah komprehensif diabetes melitus: klasifikasi, gejala, diagnosis, pencegahan, dan pengobatan," *J. Bioteknologi Biosains Indones.*, vol. 7, no. 2, pp. 304-317, 2020, DOI: 10.29122/jbbi.v7i2.4209.

[2] M. Hayati, Yusrawati, and F. Dewi, "Gaya Hidup Kafetaria dan Prediktor Preventif Diabetes Mellitus Remaja," *J. Telenursing*, vol. 5, no. 2, pp. 3314-3322, 2023, DOI: 10.31539/joting.v5i2.8018.

[3] D. W. Hestiana, "Factors Associated with Compliance in Diet Management in Outpatients with Type 2 Diabetes Mellitus in Semarang City," *J. Health Educ.*, vol. 2, no. 2, pp. 137-145, 2017, DOI: 10.15294/jhe.v2i2.14448.

[4] A. Nurmeliana, F. Octeriyedi, and Murniarni, "Factors Related to Dietary Patterns in Type II Diabetes Mellitus Patients Aged 45-55 Years in the Mentok Community Health Center Working Area, West Bangka Regency," *J. Pendidik. Tambusai*, vol. 8, no. 1, pp. 15084-15094, 2024, DOI: 10.31004/jptam.v8i1.14522.

[5] T. Mulyanto, and R. Saifudin, "The Relationship between Diet and Physical Activity Patterns and the Incidence of Diabetes Mellitus at the Tanjung Priok Health Quarantine Center," *MAHESA: Malahayati Health Stud. J.*, vol. 5, no. 10, pp. 4377-4386, 2025, DOI: 10.33024/mahesa.v5i10.19466.

[6] G. D. Ogle, F. Wang, G. A. Gregory, and J. Maniam, "Type 1 Diabetes Estimates in Children and Adults," *IDF Atlas Report*, 2022.

[7] N. W. H. M. H. Murti, B. K. Dewi, and S. Nurjanah, "Outreach and Control of the Non-Communicable Disease, Diabetes Mellitus, in Tenajar Village, Kertsemaya District, Indramayu Regency," *BEGAWE: J. Pengabdian Masyarakat*, vol. 1, no. 2, pp. 58–65, 2023, DOI: 10.62667/begawe.v1i2.22.

[8] S. Nurohmi, "Differences in Vegetable and Fruit Consumption in Normal Subjects and Patients with Type 2 Diabetes Mellitus." *Darussalam Nutr. J.*, vol. 1, no. 2, pp. 37-44, 2017, DOI: 10.21111/dnj.v1i2.1344.

[9] G. Siahaan, E. Nainggolan, and D. Lestrina, "The Relationship between Nutrient Intake and Triglyceride and Blood Glucose Levels in Vegetarians," *Indones. J. Hum. Nutr.*, vol. 2, no. 1, pp. 48–60, 2015, DOI: 10.21776/ub.ijhn.2015.002.01.5.

[10] E. Riadi and A. Prabawati, "*Research Statistics*." Yogyakarta, Indonesia: Andi Publisher, 2016.

[11] Tendriyawati, G. N. A. Wibawa, and B. Abapihi, "Poisson Regression Modeling of Factors Influencing the Occurrence of Hypertension in Kendari City," *J. Matematika, Komputasi Statistika*, vol. 3, no. 1, 2023, DOI: 10.33772/jmks.v3i1.35.

[12] G. Kholijah, N. Rarasati, and C. Sormin, "Poisson Regression Model for Estimating Covid-19 Affecting Factors on Sumatra Island," *UNEJ e-Proc.*, pp. 533-538, 2022.

[13] Y. Setyawan, K. Suryowati, and D. Octaviana. "Application of Negative Binomial Regression Analysis to Overcome the Overdispersion of Poisson Regression Model for Malnutrition Cases in Indonesia." *Parameter: J. Stat.*, vol. 2, no. 2, pp. 1-9, 2022, DOI: 10.22487/27765660.2022.v2.i2.15903.

[14] R. T. Simarmata and D. Ispriyanti, "Handling Overdispersion in Poisson Regression Model Using Negative Binomial Regression Model," *Media Statistika*, vol. 4, no. 2, pp. 95-104, 2011, DOI: 10.14710/medstat.4.2.95-104.

[15] R. Cahyandari, "Overdispersion Testing in Poisson Regression Model," *Statistika*, vol. 14, no.2, 2014.

[16] D. Sulistianingrum, "Durbin Spatial Regression Analysis for Malnutrition Cases in Toddlers in East Java Province," Doctoral Dissertation, Universitas Negeri Jakarta, 2015.

[17] T. W. Utami, "Negative Binomial Regression Analysis to Overcome Poisson Regression Overdispersion in Dengue Fever Cases," *J. Statistika Universitas Muhammadiyah Semarang*, vol. 1, no. 2, 2013, DOI: 10.26714/jsunimus.1.2.2013.%25p.