

Analysis of Environmental and Productivity Factors with the Number of Dengue Hemorrhagic Fever and Obesity Cases in Indonesia

Abu Hanifah Al Faruqy⁽¹⁾, A'yunin Sofro⁽²⁾.

^{1,2}Universitas Negeri Surabaya

Jl. Ketintang, Ketintang, Kec. Gayungan, Kota Surabaya, Jawa Timur

e-mail: ayuninsofro@unesa.ac.id⁽²⁾

ABSTRAK

Demam Berdarah Dengue (DBD) adalah penyakit yang sering terjadi di daerah tropis. Hal ini terjadi karena vektor pembawa penyakit ini adalah nyamuk *Ae. aegypti* yang habitatnya berada di lingkungan tropis. Obesitas telah menjadi masalah global bagi seluruh belahan dunia. Pasien dengan gizi obesitas mempunyai respon imunitas yang lebih kuat karena terjadi peningkatan inflamasi di dalam sirkulasi. Sehingga pembuluh darah dapat menjadi lebih lebar dari biasanya. Hal ini memicu kebocoran plasma dan memperparah penyakit DBD sehingga dapat menimbulkan SSD. Kedua penyakit ini memiliki faktor pemicu bermacam-macam seperti lingkungan dan produktivitas. Dalam penelitian ini, akan menggunakan analisis regresi linear multivariat untuk mengetahui faktor mana saja yang signifikan terhadap kedua penyakit tersebut. Analisis regresi linear menggunakan uji wilks lambda secara serentak serta parsial. Berdasarkan hasil penelitian dinyatakan bahwa ada 4 variabel prediktor yang signifikan terhadap tingkat kasus DBD dan obesitas yaitu, Indeks Kualitas Udara (x_1), produktivitas (x_2), tingkat kemiskinan (x_3) dan jumlah PUSKESMAS (x_4).

Kata kunci: DBD, Obesitas, Analisis Regresi Linear Multivariat

ABSTRACT

*Dengue Hemorrhagic Fever (DHF) is a frequently occurring disease in tropical regions. This is due to the disease vector being the *Aedes aegypti* mosquito, whose habitat is in tropical environments. Obesity has become a global issue worldwide. Patients with obesity have a stronger immune response due to increased inflammation in circulation. As a result, blood vessels may become wider than usual. This triggers plasma leakage and exacerbates DHF, potentially leading to Severe Dengue Syndrome (SDS). Both of these diseases have various triggering factors such as environmental and productivity-related aspects. In this study, multivariate linear regression analysis will be employed to identify which factors are significant for both diseases. The linear regression analysis will use simultaneous and partial Wilks' Lambda tests. Based on the research findings, it is stated that there are four predictor variables significant for the levels of DHF and obesity, namely, Air Quality Index (x_1), productivity (x_2), poverty rate (x_3), and the number of health centers (x_4).*

Keywords: Regression Linear Multivariat; Analysis; DHS; Obesit

INTRODUCTION

Dengue Hemorrhagic Fever (DHF) is a disease that frequently occurs in tropical areas and become a global concern [1], [2]. This happens because the vector carrying this disease is the *Aedes aegypti* mosquito, which habitat is in tropical environments. According to the World Health Organization, DHF is estimated to have 50 to 100 million cases of infection, with 22,000 deaths occurring. The most severe cases of DHF occurred in 2019, with America reporting 3.1 million cases and in Asia itself, reports from Bangladesh (101,000), the Philippines (420,000), Vietnam (320,000), and Malaysia (131,000) [3]. In Indonesia, DHF cases increased from 65,602 cases in 2018 to 138,127 cases in 2019. This indicates that the possibility of DHF occurring in Indonesia is still very high [4].

Obesity has become a global issue for the entire world [5]. Data shows that in the United States in 2007, 2.2 percent of men and 35.5 percent of women had a Body Mass Index (BMI) above normal [6]. In Indonesia, according to the Ministry of Health (KEMENKES), obesity is one of the diseases that will be targeted for eradication.

Given this global trend, obesity is often associated with an increased risk of dengue shock syndrome (DSS) or critical phase dengue hemorrhagic fever (DHF) cases. The prevalence of obesity in individuals over 18 years old reaches 12 percent [7]. Patients with obesity have a stronger immune response due to increased inflammation in the circulation. This can cause blood vessels to become wider than usual. This triggers plasma leakage and worsens the condition of DHF, potentially leading to DSS [3].

So, people with obesity are more susceptible to dengue fever than other people. This could be a serious condition because if obesity individuals infected with dengue fever, the obesity patient will increase the severity of dengue fever [8]. With the high number of obesity and dengue fever cases in the whole world, it is important, especially for people with obesity to take preventive steps.

By deeper understanding of the pattern of both diseases is key to control and prevention strategy [2]. To prevent this from happening, preventive measures can be taken by controlling the causes of both dengue fever and obesity. Environmental factors, including the cleanliness of the environment, weather conditions, and pollution influence the spread of dengue hemorrhagic fever (DHF) itself [9], [10]. Unlike DHF, obesity is not a contagious disease but can occur when a person does not maintain a healthy lifestyle, one aspect of which is the frequency of physical activity [11], [12], [13], [14].

Multivariate linear regression is a regression analysis that involves more than one response variable (y) and more than one predictor variable (x) that are correlated with each other [14], [15]. Therefore, to get to know which causes that possibly to be controlled, this multivariate linear regression analysis can be used as a tool to determine the extent to which environmental factors and productivity affect DHF and obesity. Hence, this study is titled "Analysis of Environmental and Productivity Factors with the Number of Dengue Hemorrhagic Fever and Obesity Cases in Indonesia."

METHOD

The research method used is multivariate regression analysis, with data collection stages from the official websites of the Ministry of Health, the Central Statistics Agency, and the Directorate of Air Pollution Control, Directorate General of Pollution and Environmental Damage

Control. The collected data can then be processed using multivariate regression analysis with the help of R Studio, and the research steps are as follows:

1. Determine the descriptive statistical analysis of the research data.
2. Determine the estimation of multivariate linear regression parameters.
3. Perform simultaneous and partial significance testing of the parameters using the Wilks' Test.
4. Determine the multivariate linear regression model.
5. Determine the simple linear regression model.
6. Conduct the Akaike Information Criterion (AIC) test on the simple and multivariate linear regression models.
7. Draw conclusions and make recommendations.

The response variables used in this study are as follows :

- y_1 : Dengue Hamorrhagic Fever (DHF) case rate
- y_2 : Obesity prevalence rate

There are 6 predictor variables in this study are as follows:

- x_1 : Air Quality Index (AQI)
- x_2 : Work productivity level
- x_3 : Poverty rate
- x_4 : Number of community health centers
- x_5 : Unemployment rate
- x_6 : Level of heavy activity in the past week

In this study, environmental factors are derived from variables x_1 dan x_4 . The chosen variables representing environmental factors because of the dengue fever can be controlled by the air quality and treatments of dengue disease case in health center. While productivity factors are derived from variables $x_2, x_3, x_5,$ dan x_6 . Therefore, these variables represent productivity because obesity is a disease that related to lifestyle, so these variables were chosen.

RESULT AND DISCUSSION

There are 2 response variables consisting of y_1 and y_2 dan 6 predictor variables, namely, $x_1, x_2, x_3, x_4, x_5,$ and x_6 which will be used in this study. Descriptive statistics are an initial statistical analysis to understand basic information in the data [17]. Here is the descriptive statistics for the entire research dataset

Table 1. Descriptive Statistics of Study.

Variable	Minimum	Maximum	Mean (Average)
y_1	2,9	80,90	28,31
y_2	13	42	29,76
x_1	68,06	95,79	89,18
x_2	24,94	400	94,5
x_3	4,6	27,4	10,43
x_4	0,29	7,16	1,59

x_5	2,27	7,53	4,61
x_6	40,66	75,01	53,09

Multivariate Linear Regression Analysis

Multivariate linear regression analysis is a linear regression analysis that involves more than one or more response variables (y) and predictor variables (x) [18]. Subsequently, from the response variables and predictor variables, a multivariate linear regression model with i responses can be formed as follows.

$$\begin{aligned}
 y_1 &= \beta_{01} + \beta_{11}x_1 + \beta_{21}x_2 + \dots + \beta_{n1}x_n + \epsilon_1 \\
 y_2 &= \beta_{02} + \beta_{12}x_1 + \beta_{22}x_2 + \dots + \beta_{n2}x_n + \epsilon_2 \\
 &\vdots \\
 y_i &= \beta_{0i} + \beta_{1i}x_1 + \beta_{2i}x_2 + \dots + \beta_{ni}x_n + \epsilon_i
 \end{aligned}
 \tag{1}$$

In determining regression analysis, there are several steps to follow. Firstly, estimating the parameters.

Estimated Parameter Results

One of the main objectives of regression analysis is to express the predicted value of the response variable through the values of the predictor variables from the regression equation. Before determining the regression equation, the regression coefficients, denoted as β . $\bar{\beta}$ can be obtained from the following equation [19] :

$$\bar{\beta} = (X^T X)^{-1} X^T Y
 \tag{2}$$

Where,

X : Matrix of predictor variables $n \times (n + 1)$

Y : Matrix of response variable n

$\bar{\beta}$: Matrix of regression parameters

Thus, from the construction of parameter estimation for each variable, the following is obtained :

Table 2. Estimated parameter results each predictor variables regarding the response variables

Parameter	y_1	y_2
x_0	-26,7373	-72,3188
x_1	0,1101	0,9716
x_2	0,0089	0,0284
x_3	3,5636	0,4109
x_4	2,5688	-1,4411
x_5	-0,8525	0,9694
x_6	0,1328	0,1182

Based on the estimated parameter results above, a temporary multivariate regression model for y_1

and y_2 can be formulated as follows:

$$\begin{aligned}
 y_1 &= -26,7373 + 0,1101x_1 + 0,0089x_2 + 3,5636x_3 + \\
 &\quad 2,5688x_4 - 0,8525x_5 + 0,1328x_6 \\
 y_2 &= -72,3188 + 0,9716x_1 + 0,0284x_2 + 0,4109x_3 - \\
 &\quad 1,4411x_4 + 0,9694x_5 + 0,1182x_6
 \end{aligned}
 \tag{3}$$

Parameter Significance Testing

Significance testing is used to determine whether all parameters have a significant effect or not. The significance level can be observed; if all parameters have a value of zero, then there is no influence from the predictor variables on the response variable. Conversely, if the parameter values are not zero, there is an influence from the predictor variables on the response variable [19]. In multivariate regression analysis, there are several tests: simultaneous test and partial test.

Simultaneous Testing

Simultaneous testing is used to determine whether all parameters collectively have a significant effect on the model. Simultaneous testing is performed using Wilk’s Lambda test with the following hypotheses:

$$H_0 : \bar{\beta}_{11} = \bar{\beta}_{12} = \dots = \bar{\beta}_{in} = 0 \text{ (model is not significant)}$$

$$H_1 : \text{paling sedikit ada satu } \bar{\beta}_{in} \neq 0 \text{ (model significant)}$$

The computed Wilk’s Lambda test result is $\lambda_{calculate} = 0.0156$ with a $p - value = 2.22 \times 10^{-16}$. Thus, it can be concluded that the $p - value < \alpha = 0.05$, Therefore, the decision is to reject H_0 and the model is significant collectively.

Partial Testing

Using the same hypothesis, partial testing will be conducted for each parameter. The results of partial testing are as follows:

Table 3. Variable analysis with Wilks value and p-value result

Variable	Nilai Wilks	p-value
x_1	0,4912	$9,636 \times 10^{-5}$
x_2	0,7214	0,0143
x_3	0,0272	$2,2 \times 10^{-16}$
x_4	0,7694	0,0331
x_5	0,9203	0,3399
x_6	0,9201	0,3389

Based on Table 3., the predictor variables x_1, x_2, x_3 and x_4 have p-values for each variables less than $\alpha(0,05)$, thus it can be concluded that x_1, x_2, x_3 and x_4 significantly influence the response variables y_1 and y_2 . Conversely, the predictor variables x_5 and x_6 have p-values greater than $\alpha(0,005)$ indicating that x_5 and x_6 are not significant for the response variables y_1 and y_2 .

Parameter estimation will be performed with significant variables, resulting in the following:

Table 4. Significant estimated parameters

Variable	y_1	y_2
β_0	-30,1027	-52,0933
β_1	0,1951	0,8573
β_2	0,0074	0,0308
β_3	3,5318	0,4865
β_4	2,1728	-1,6222

Multivariate Regression Model

Based on the partial Wilks' test results, there are 4 predictor variables that have a significant influence on the response variable. These predictor variables are x_1, x_2, x_3 and x_4 . Therefore, the multivariate linear regression model can be written as follows:

$$\begin{aligned}
 y_1 &= -30,1027 + 0,1951x_1 + 0,0074x_2 + 3,5318x_3 + 2,1728x_4 \\
 y_2 &= -52,0933 + 0,8573x_1 + 0,0308x_2 + 0,4865x_3 - 1,6222x_4
 \end{aligned}
 \tag{4}$$

From this multivariate linear regression model, the interpretations are as follows:

- For every one percent increase in the Air Quality Index (AQI), the number of DHF cases will increase by 0.1951. For every one percent increase in work productivity level, the number of DHF cases will increase by 0.0074. For every one percent increase in the poverty rate, the number of DHF cases will increase by 3.5318, and for every one percent increase in the number of community health centers, the number of DHF cases will increase by 2.1728.
- For every one percent increase in the Air Quality Index (AQI), the obesity prevalence rate will increase by 0.8573. For every one percent increase in work productivity level, the obesity prevalence rate will increase by 0.0308. For every one percent increase in the poverty rate, the obesity prevalence rate will increase by 0.4865, and for every one percent increase in the number of community health centers, the obesity prevalence rate will decrease by 1.6222.

Simple Linear Regression Analysis

Linear regression analysis is a method used to establish a functional relationship between two variables. Typically, this analysis involves one response variable and one predictor variable. The regression model can be written as follows [20].

$$y = a + bx. \tag{5}$$

In this study, the regression of the response variable on each predictor variable can be observed as follows:

Table 5. Estimated parameters variables that significant to y_1

Variable	y_1
a	-30,1027
x_1	0,1951

x_2	0,0074
x_3	3,5318
x_4	2,1728

According to simple linear regression result, can constructed regression model as follows:

$$y_1 = -30,1027 + 0,1951x_1 + 0,0074x_2 + 3,5318x_3 + 2,1728x_4 \quad (6)$$

Table 6. Estimated parameters variables that significant to y_2

Variable	y_2
a	-52,0933
x_1	0,8573
x_2	0,0308
x_3	0,4865
x_4	-1,6222

According to simple linear regression result, can constructed regression model as follows:

$$y_2 = -52,0933 + 0,8573x_1 + 0,0308x_2 + 0,4865x_3 - 1,6222x_4 \quad (7)$$

Akaike Information Criterion (AIC)

AIC (Akaike Information Criterion) is used to evaluate how well a model fits the data. The AIC value chosen is the smallest among other AIC values [21,22,23]. In this study, the AIC values of the simple linear regression model and the multivariate linear regression model are compared. The AIC values of both models can be seen as follows:

Table 7. Comparison of AIC values between the simple linear regression model and the multivariate linear regression model

Model	Nilai AIC
Regresi Linear Sederhana	219,55
Regresi Linear Multivariat	145,86

According AIC values, the model with the smallest AIC value is considered the best model. Therefore, the model from multivariate linear regression is selected as the chosen regression model in this study.

CONCLUSION

Based on the results of the Wilks' Lambda test, there are 4 predictor variables that significantly influence the response variable, namely x_1 (Air Quality Index), x_2 (Work productivity), x_3 (Poverty rate), and x_4 (Number of community health centers). From these four variables, a linear regression model can be derived. In this study, simple linear regression analysis was also conducted on each predictor variable against each response variable. Based on the results

of both simple linear regression and multivariate linear regression analyses, two different AIC values were obtained. Therefore, the model with the smallest AIC value will be chosen as the best model in this study. Thus, the multivariate linear regression model is selected as the best model, with the following model:

$$\begin{aligned} y_1 &= -30,1027 + 0,1951x_1 + 0,0074x_2 + 3,5318x_3 + 2,1728x_4 \\ y_2 &= -52,0933 + 0,8573x_1 + 0,0308x_2 + 0,4865x_3 - 1,6222x_4 \end{aligned} \quad (8)$$

REFERENCE

- [1] “World Mosquito Program,” Dengue. [Online]. Available: <https://www.worldmosquitoprogram.org/en/learn/mosquito-borne-diseases/dengue#>
- [2] L. P. S. Pratiwi, “Binary Logistic Regression Analysis on the Spread of Dengue Fever in Bali Province,” vol. 17, no. 1, 2024.
- [3] R. A. A. Naiem, R. Rompies, and S. N. N. Tatura, “Hubungan antara Status Nutrisi dengan Tingkat Keparahan Infeksi Dengue pada Pasien Anak di RSUP Prof. Dr. R. D. Kandou, Manado, Indonesia,” *eCL*, vol. 11, no. 1, Oct. 2022, doi: 10.35790/ecl.v11i1.37887.
- [4] M. Arisanti and N. H. Suryaningtyas, “KEJADIAN DEMAM BERDARAH DENGUE (DBD) DI INDONESIA TAHUN 2010-2019,” *J.SPIRAKEL*, vol. 13, no. 1, pp. 34–41, Dec. 2021, doi: 10.22435/spirakel.v13i1.5439.
- [5] S. Mugianti, T. Setijaningsih, and K. Fransiska, “Perkembangan Motorik Balita Obesitas Usia 3-5 Tahun,” *JNK JOURNAL*, vol. 5, no. 1, pp. 046–052, Apr. 2018, doi: 10.26699/jnk.v5i1.ART.p046-052.
- [6] Ni Putu Indah Kartika Putri and I Made Gede Dwi Lingga Utama, “HUBUNGAN OBESITAS DENGAN KEJADIAN SINDROM SYOK DENGUE PADA ANAK,” *JURNAL MEDIKA UDAYA*, 10 2020.
- [7] KEMENKES RI, *Laporan Kinerja Kementerian Kesehatan Tahun 2022*. 2022.
- [8] Naseem S. Miller, “Dengue: A primer and research roundup,” Aug. 2024, [Online]. Available: <https://journalistsresource.org/home/dengue-research-roundup/>
- [9] L. L. Ayun and E. T. Pawenang, “Hubungan antara Faktor Lingkungan Fisik dan Perilaku dengan Kejadian Demam Berdarah Dengue (DBD) Di Wilayah Kerja Puskesmas Sekaran, Kecamatan Gunungpati, Kota Semarang,” 2017.
- [10] H. Husin, A. Ramon, W. Angraini, N. Wati, and P. Anugrah, “ANALISIS FAKTOR RISIKO YANG BERHUBUNGAN DENGAN KEJADIAN DEMAM BERDARAH DENGUE (DBD)”.
- [11] D. D. Amanda and S. Martini, “Relationship Of Characteristic and Status of Central Obesity with The Prevalence Of Hypertension,” *JBE*, vol. 6, no. 1, p. 57, Mar. 2018, doi: 10.20473/jbe.V6I12018.57-66.
- [12] W. Widiyanti and Z. Tafal, “Aktivitas Fisik, Stres, dan Obesitas pada Pegawai Negeri Sipil,” *Kesmas: National Public Health Journal*, p. 325, Feb. 2014, doi: 10.21109/kesmas.v0i0.374.
- [13] I. Trisna and S. Hamid, “FAKTOR-FAKTOR YANG BERHUBUNGAN DENGAN OBESITAS SENTRAL PADA WANITADEWASA (30-50 TAHUN) DI KECAMATAN LUBUK SIKAPING TAHUN 2008,” *JKMA*, vol. 3, no. 2, pp. 68–71, Mar. 2009, doi: 10.24893/jkma.v3i2.62.
- [14] M. Wirastuti and A. Sofro, “Analisis Longitudinal pada Data Pasien Diabetes Melitus,” *J Statistika*, vol. 12, no. 1, pp. 13–19, Jul. 2019, doi: 10.36456/jstat.vol12.no1.a1993.
- [15] S. Sukirman, “Analisis Cloud Computing Untuk Penyimpanan Dokumen Terhadap Proses Pembelajaran Menggunakan Algoritma Regresi Linear Berganda (Studi Kasus : SMA Chandra

- Kusuma Jakarta Utara),” *Jurnal.ilmiah.informatika*, vol. 7, no. 1, pp. 1–12, Jun. 2022, doi: 10.35316/jimi.v7i1.1-12.
- [16] P. R. Arum, R. P. Gautama, I. Fitriani, and F. Naza, “Identifying Factors that Influence Life Expectancy in Central Java Using Spatial Regression Models,” vol. 16, no. 2, 2023.
- [17] L. D. Martias, “STATISTIKA DESKRIPTIF SEBAGAI KUMPULAN INFORMASI,” *FHRS*, vol. 16, no. 1, p. 40, Jun. 2021, doi: 10.14421/fhrs.2021.161.40-59.
- [18] Richard A J et al and Dean W, *Applied multivariate statistical analysis*, 5th ed. 2002.
- [19] Alvin C. Rencher, *Multivariate Statistical Inference and Applications*. A Wiley-Interscience Publication, 1998.
- [20] N. Suhandi, E. A. K. Putri, and S. Agnisa, “Analisis Pengaruh Jumlah Penduduk terhadap Jumlah Kemiskinan Menggunakan Metode Regresi Linear di Kota Palembang,” vol. 09, no. 2, 2018.
- [21] R. Bevans, “Akaike Information Criterion | When & How to Use It (Example),” *Scibbr*, Jun. 2023, [Online]. Available: <https://www.scribbr.com/statistics/akaike-information-criterion/>
- [22] A. Salsabila, Erfiani, Indahwati, A. . Fitrianto, and M. A. . Aliu, “Association of Poverty Categories, Educational Characteristics, and Area of Residence in Indonesia Using a Three-Way Log-Linear Model,” *J Statistika*, vol. 17, no. 1, pp. 624–634, Jul. 2024.
- [23] G. Anuraga, A. Indrasetianingsih, and M. Athoillah, “Pelatihan pengujian hipotesis statistika dasar dengan software r,” *BUDIMAS: Jurnal Pengabdian Masyarakat*, vol. 3, no. 2, pp. 327–334, 2021.