ENTHUSIASTIC
INTERNATIONAL JOURNAL OF APPLIED STATISTICS AND DATA SCIENCE
Volume 5, Issue 2, October 2025, pp. 166-177

Log-Linear Analysis of the Association among Hematological Variables in Dengue Hemorrhagic Fever Cases

Miftahul Irfan [a,1,*], Ma'rufah Hayati [a,2], Nora Madonna [a,3], Wardhani Utami Dewi [b,4]

[a] Institut Teknologi Sumatera, Jl. Terusan Ryacudu, Way Huwi, Kec. Jati Agung, Lampung Selatan, Lampung and 35365, Indonesia
[b] Universitas Muhammadiyah Metro, Jalan Ki Hajar Dewantara, Metro Timur, Kota Metro, Lampung
[1] miftahul.irfan@at.itera.ac.id *; [2] marufah.mt@at.itera.ac.id; [3] nora.madonna@at.itera.ac.id, [4] wardhaniutamidewi@ummetro.ac.id
* Corresponding author

ARTICLE INFO

Article history
Received: September 5, 2025
Revised: October 25, 2025
Accepted: October 31, 2025

Keywords
Log-Linear Modeling
Leukocyte Count
Hemoglobin Level
Hematocrit Value
Interaction Pattern

ABSTRACT

Health data are often analyzed in their continuous form through approaches such as linear, logistic, or survival models. In this study, hematological variables were dichotomized based on established clinical cut-offs to enable log-linear analysis of associations among categorical variables, acknowledging the potential loss of information from this transformation. A log-linear model was applied to evaluate independence, dependence, and interaction patterns among leukocyte, hemoglobin, and hematocrit categories in a dengue hemorrhagic fever (DHF) patient dataset. Previous analyses using survival models identified these variables as factors associated with recovery rates; however, these models did not capture their interaction structure. Log-linear analysis was therefore employed to examine these associations more comprehensively. The best-fitting model was identified as $L + Hb + H + L * H + Hb * H$, which included two-factor interactions between leukocyte–hematocrit and hemoglobin–hematocrit. This model demonstrated a good fit (Pearson $X^2 = 1.773$, $df = 2$, $p = 0.412$), including a three-factor interaction resulted in a saturated model ($df = 0$) and did not improve model performance. These findings highlight significant interaction patterns among hematological variables in DHF patients, providing a more detailed understanding of their joint associations.

1. Introduction

Statistical research generally uses numerical data, whereas research utilizing categorical data remain relatively limited. This limitation poses a problem for data analysis methods because the methods commonly used for categorical data differ from those applied in numerical data. In fact, there are many variables that use categorical data such as gender and age (children, adults, elderly). Health data utilized in research should generally be classified as research, according to [1]–[4]. This study also used health data in the form of categories. The categories used were normal or abnormal on the variables of leukocytes, hemoglobin, and hematocrit. These variables were categorized based on established clinical cut-offs to ensure clinically meaningful groupings. This approach also aligns with the methodological requirement of constructing a contingency table for log-linear analysis, which is designed for categorical data.

p-ISSN 2798-253X
e-ISSN 2798-3153

Logistic regression is a commonly used for analyzing categorical data when there is a clearly defined dependent variable. In contrast, log-linear modeling treats all variables symmetrically and focuses on modeling cell counts in contingency tables to examine associations and interaction structures. However, the variables used are dichotomous variables which only consist of two values. Therefore, the log-linear method is proposed. Log linear analysis can be used for categorical data analysis with two variables, three variables or even more than three variables. According to [5] the log-linear model is a method used to analyze categorical data on a nominal scale. According to [6], category data analysis using the linear log method is the development of a two-way contingency table. Categorical variables were analyzed using the natural logarithm of the cell frequencies in the contingency table. The contingency table analysis itself is used to see whether there is a relationship between one variable and another.

Although the log-linear model can be used to analyze the relationship between two categorical variables in a contingency table, the log-linear model is also usually used to evaluate a multi-way contingency table involving three or more variables [7]. A complete three-variable model is a model with all variables plus the effects of two-factor interactions and three-factor interactions. In log-linear models, no variable is designated as the dependent variable; all variables are treated symmetrically to examine patterns of association and interaction among categorical variables [8]. The basic purpose of this model is to describe patterns of association and interaction between different levels of categorical variables [9]. Hematological parameters such as leukocyte count, hemoglobin concentration, and hematocrit levels are among the key clinical indicators used to monitor disease progression and severity in dengue patients. Recent studies have reported that abnormal hematological profiles are strongly associated with disease status and clinical outcomes in dengue infection [10]–[12]. These parameters often show characteristic patterns, including leukopenia, hemoconcentration, and hemoglobin fluctuations during the acute phase.

Dengue hemorrhagic fever (DHF) is one of the diseases that consistently becomes a major concern for both the government and the community during the rainy season. It is a disease that can spread rapidly and may causes death within a short period. One modeling estimate has suggested 390 million dengue virus infections per year (95% credible interval: 284–528 million), with 96 million (67-136 million) cases manifesting clinically across various disease severity [13]. Another study on the prevalence of DHF has estimated that 3.9 billion people are at risk of being infected with the dengue virus. Although the risk of infection is present in 129 countries [14], 70% of the disease burden occurs in Asia [13]. The World Health Organization (WHO) reported that the number of dengue cases increased more than eightfold over the last two decades, rising from 505,430 cases in 2000, to more than 2.4 million in 2010, and 4.2 million in 2019 [15]. Deaths reported between 2000 and 2015 increased from 960 to 4,032.

A previous study applied a survival analysis to DHF patients at Haji Hospital Surabaya using the Weibull regression model [16]. On the other hand, between January to March 2020, Lampung province ranked first in DHF cases, reporting 3,431 cases and 13 deaths. Another study employing survival analysis in DHF patients reported that leukocytes, hemoglobin, and hematocrit influenced the recovery rate of DHF patients in Lampung province [17]. Therefore, using the same dataset, the analysis focused on assessing associations and interaction patterns among leukocytes, hemoglobin, and hematocrit categories within the DHF patient sample using a log-linear modeling framework. The best models obtained were also compared. The focus of this study was to examine how the interaction among variables affected the model, as the variables used were variables that theoretically influenced DHF disease. In addition, a comparison of models was carried out to determine which model yielded better results, the one with interaction terms or the one without.

2. The Proposed Method

2.1. Categorical Data

Categorical data are data containing categorical variables as well as data in the form of frequency of observations. They will be easier to analyze if they are presented in the form of a contingency

p-ISSN 2798-253X
e-ISSN 2798-3153

table. Categorical data are derived by classifying all individual samples into one or more categorical variables simultaneously. Age and gender are the examples of categorical data. Age can be categorized as children, adolescents, adults, and the elderly, whereas gender while gender can be categorized. Another example is the results of laboratory measurements, including leukocyte count, hemoglobin concentration, and hematocrit percentage, as they can be categorized as normal and abnormal based on standard clinical reference thresholds. Based on the clinical laboratory reference standards of Abdul Moeloek Hospital and WHO dengue clinical guidelines [18], [19], leukocyte counts are classified as abnormal if < 4.0 × 10^9/L. Meanwhile, hemoglobin levels are classified as abnormal if < 13 g/dL for males and < 12 g/dL for females. Hematocrit values are considered abnormal if below 40–50% (males) and 36–44% (females) or above the upper reference limit indicating hemoconcentration.

2.2. 3D Log-Linear Model

The log-linear model is a statistical approach used to analyze the relationship among categorical variables by modeling the natural logarithm of expected cell frequencies in a contingency table as a linear function of parameters representing main and interaction effects. Unlike regression models that emphasize prediction, log-linear models treat all variables symmetrically and focus on examining associations or dependencies among categorical variables. In this study, hierarchical models represented as $[L][Hb][H]$ (independence model), $[L*H][Hb*H]$ (two-factor interaction), and $[L*Hb*H]$ (three-way interaction or saturated model) were evaluated. Model fit was assessed using Pearson chi-square (X^2) and likelihood ratio (G^2) statistics. In the case of three categorical variables, a three-way contingency table with *i* rows, *j* columns, and *k* layers was employed, allowing researchers to investigate not only the main effects but also two-way and three-way interactions. This framework enables the identification of independence, conditional independence, and interaction structures, making the log-linear model a powerful tool in understanding complex relationships within multidimensional categorical data, particularly in fields such as social sciences, epidemiology, actuarial science, and marketing research [20].

3. Method

This section provides a detailed explanation of the research method employed as well as the variables used in the study. The chosen research method was designed to ensure that the processes of data collection, processing, and analysis were conducted systematically and in accordance with the research objectives. Furthermore, the identification of research variables represented an essential step, as these variables served as the foundation for constructing the analytical model and addressing the research questions. Thus, the explanation of both the research method and the selected variables will offer a clear understanding of the analytical approach undertaken and its relevance to the research topic.

3.1. Data and Variables

The data used in this study were DHF data obtained from Abdul Moeloek Hospital, Lampung Province. The data collected were data on 59 DHF patients from January 2019 to January 2020. The variables used were leukocytes, hemoglobin and hematocrit variables. Each variable was categorized as either normal or abnormal.

Table 1. Research Data

Leukocytes	Hemoglobin	Hematocrit	
		Normal	*Abnormal*
Normal	Normal	10	16
	Abnormal	0	11
Abnormal	Normal	12	4
	Abnormal	1	5

From Table 1 exhibits the data employed Leukocyte, hemoglobin, and hematocrit values were categorized according to standard clinical reference thresholds. Leukocyte counts are considered abnormal if < 4.0 × 10^9/L, hemoglobin levels are considered abnormal if < 13 g/dL for males or < 12

p-ISSN 2798-253X
e-ISSN 2798-3153

g/dL for females, and hematocrit values are considered abnormal if outside the reference range of 40–50% for males or 36–44% for females [19]. This categorization approach is consistent with recent dengue studies that have identified abnormal leukocyte counts, hemoglobin levels, and hematocrit values as key markers of disease severity [21]. These thresholds were based on WHO dengue clinical guidelines and the laboratory reference manual of Abdul Moeloek Hospital [19].

3.2. Research Stage

This research was conducted through several structured stages to ensure systematic and reliable results. The first stage was descriptive statistics, where the percentage of each variable was presented to provide an overview of the data distribution. Bar charts were used to visually illustrate the patterns among variables, and the chi-square test was applied to assess whether significant associations existed between them. The second stage was modeling and model selection. In this step, several log-linear models were constructed and evaluated based on their goodness-of-fit statistics. This procedure aimed to identify the most appropriate model that accurately represents the relationships among the variables.

The third stage involved parameter estimation, where parameter values were derived from the selected model to determine which main and interaction effects were statistically significant. Finally, the conclusion stage summarized the analytical findings and provided interpretations of the observed relationships among leukocyte, hemoglobin, and hematocrit variables in DHF patients.

3.3. Independent Model

Independent model means a model that is formed without any interaction of the variables which $u_{12(ij)} = u_{13(ik)} = u_{23(jk)} = u_{123(ijk)}$ [6] has the form:

$$\text{Log } m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} \quad (1)$$

where m_{ijk} denotes the expected frequency in each *ijk* cell in the model, u denotes the average general effect, $u_{1(i)}$ denotes the main effect of the 1st category *i,* $u_{2(j)}$ denotes main effect of 2nd category, and $u_{3(k)}$ denotes main effect of the 3rd category *k.*

If all the variables are mutually independent, the expected frequency estimator for each cell is (20).

$$\hat{m}_{ijk} = \frac{(n_{i..})(n_{.j.})(n_{..k})}{n^2} \quad (2)$$

with:

$$n_{i..} = \sum_{j=1}^{J} \sum_{k=1}^{K} n_{ijk}$$

$$n_{.j.} = \sum_{i=1}^{I} \sum_{k=1}^{K} n_{ijk}$$

$$n_{..k} = \sum_{i=1}^{I} \sum_{j=1}^{J} n_{ijk}$$

$n_{...} = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} n_{ijk}$.

3.4. One-Variable Interaction Model

The two-variable interaction model indicates that the presence of interaction allows dependency between variables: between variable 1 and 2 with variable 3 present, variables 1 and 3 with variable 2 present, or variables 2 and 3 with variable 1 present. If, among three variables, there is only one interaction that occurs between two variables, then the model is as follows:

$$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} \quad (3)$$

where $u_{12(ij)}$ is the interaction between two variables, which can be an interaction between 1 and 2, 1 and 3, or 2 and 3. The expected frequency estimator for each cell is as in (4).

p-ISSN 2798-253X
e-ISSN 2798-3153

$$\widehat{m}_{ijk} = \frac{(n_{ij.})(n_{..k})}{n_{...}}. \quad (4)$$

3.5. Model Without Interaction of Three Variables

The model without the interaction of three variables is a model consisting of three variables where the three variables are included in the model and each variable interacts in two directions. If there is an interaction between the three variables, the model is illustrated as in (5):

$$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)} \quad (5)$$

where $u_{123(ijk)} = 0$. The expected frequency estimator is the following.

$$\widehat{m}_{ijk} = \frac{(n_{ij.})(n_{i.k})(n_{jk})}{(n_{i..})(n_{.j.})(n_{..k})}. \quad (6)$$

3.6. Saturated Model

The last possible model is a saturated model, or a saturated model is a model that contains all possible parameters, and other parameters cannot be entered. This model can also be called a full model or full model because it consists of three variables with a two-way interaction of each variable and there is also an interaction of three variables [22]. The equation of the model is as follows:

$$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)} + u_{123(ijk)} \quad (7)$$

with a frequency estimator the expectations are as follows:

$$\widehat{m}_{ijk} = n_{ijk}. \quad (8)$$

In the log-linear model, the expected cell frequency $\widehat{m}_{ijk}$ is modeled as the exponential of a linear combination of parameters representing main effects and interactions. The parameterization is:

$$\log(\widehat{m}_{ijk}) = \mu + \lambda_i^L + \lambda_j^{Hb} + \lambda_k^H + \lambda_{ij}^{L\,H} + \lambda_{ik}^{L\,Hb} + \lambda_{jk}^{Hb\,H} + \lambda_{ijk}^{L\,Hb\,H} \quad (9)$$

where μ is the overall mean effect; λ_i^L, λ_j^{Hb}, and λ_k^H represent the main effects of each variable; and λ_{ij}^{LH}, λ_{ik}^{LHb}, λ_{jk}^{HbH}, and λ_{ijk}^{LHbH} represent the two-way and three-way interaction effects, respectively. To ensure identifiability, the constraints are imposed:

$\sum_i \lambda_i^L = \sum_j \lambda_j^{Hb} = \sum_k \lambda_k^H$, and similar zero-sum constraints for the interaction terms.

3.7. Goodness of Fit Test

Goodness of fit statistics is used compare or determine whether there is a distance between observations. To test the hypothesis in each model, the Person chi-square test was used (χ^2) by using the equation:

$$\chi^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(n_{ij} - \widehat{m}_{ij})^2}{\widehat{m}_{ij}} \quad (10)$$

or can also use the likelihood ratio test (G^2) with the (10):

$$G^2 = 2 \sum_{i=1}^{I} \sum_{j=1}^{J} n_{ij} \log\left(\frac{n_{ij}}{\widehat{m}_{ij}}\right). \quad (11)$$

3.8. Model Selection

The stepwise selection procedure combined forward inclusion and backward elimination, with model terms entered or removed based on the likelihood ratio chi-square (ΔG^2) test. Terms were considered for entry if $p < 0.05$ and removed if $p > 0.05$. Hierarchical structure was strictly preserved throughout the stepwise process, ensuring that higher-order interactions were retained only when their corresponding lower-order terms were included. If an independent variable has entered one stage, it may be excluded in the subsequent stage as it is no longer potential compared to other variables that enter the model.

4. Results and Discussion

This section presents descriptive statistics of the data used and also the comparison of the best models that are feasible to use.

4.1. Descriptive Statistics

Using descriptive statistics, the percentage of each variable can be examined to see whether there is a relationship between each variable.

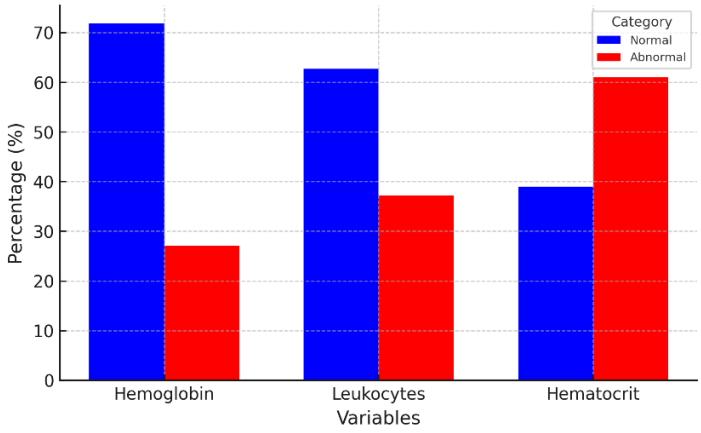

Fig. 1 Percentage of each variable.

From Fig. 1, it can be seen that the normal hemoglobin level has the largest percentage of 71.88%, meaning that the majority of patients have normal hemoglobin levels, whereas 27.12% patients have abnormal hemoglobin. While for leukocytes, the majority of patients also had normal leukocyte levels with a percentage of 62.71% and 37.21% had abnormal leukocyte levels. However, this is inversely proportional to the hematocrit level. Based on Fig. 1, it can be seen that the majority of DHF patients at RSUD Abdul Moeloek Lampung have abnormal hematocrit levels with a percentage of 61%, while those that are not normal have a percentage of 39%. Based on this description, it seems that there is an influence of the three variables on the number of DHF patients in Abdul Moeloek Hospital, Lampung.

4.1.1. 4.1.1 Relationship Between Leukocyte and Hematocrit

Fig. 2 presents a bar chart illustrating the relationship between leukocyte levels and hematocrit status among DHF patients. This visualization provides an initial overview of how variations in leukocyte counts correspond to differences in hematocrit levels, serving as a descriptive basis before further statistical analysis:

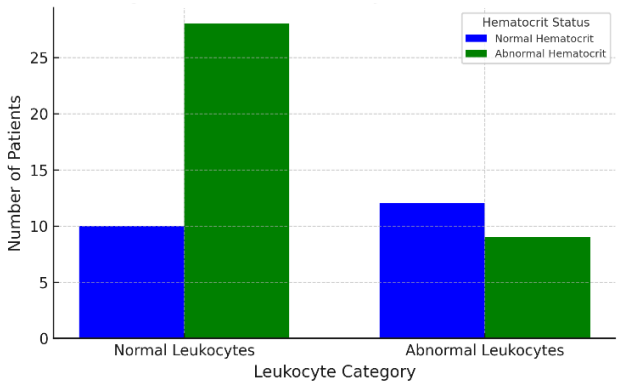

Fig. 2 Bar chart of leukocytes and hematocrit.

Among patients with normal leukocyte levels, the proportion of abnormal hematocrit levels was higher than that of normal hematocrit levels, indicating that most patients with normal leukocyte

counts tended to exhibit abnormal hematocrit levels. Conversely, among patients with abnormal leukocyte levels, the proportion of normal hematocrit levels was greater than the abnormal category, suggesting a different distribution pattern. These findings provide an initial of possible association between leukocyte and hematocrit levels. Consequently, the null hypothesis (H_0) assuming no relationship between the two variables tends to be rejected, implying the presence of a significant association.

Table 2. Chi-Square Tests 1

Test	Value	df	asymp. Sig. (2-sided)
Pearson Chi-Square	5.963a	1	.015
Likelihood Ratio	5.955	1	.015
Linear-by-Linear Association	5.862	1	.015
N of Valid Cases	59		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 8.58.
b. Computed only for a 2x2 table

Based on Table 2, the chi-square test yielded a value of 5.963 with 1 degree of freedom and a p-value of 0.015, which is below the 5% significance level ($\alpha = 0.05$). Since the p-value is less than the predetermined significance threshold, the null hypothesis (H_0) of no association is rejected. Thus, it can be concluded that there is a statistically significant relationship between leukocyte levels and hematocrit levels. This statistical evidence reinforces the descriptive findings from the Fig. 2, which suggested the presence of an association between the two variables.

4.1.2. Relationship between Hemoglobin and Hematocrit

Fig. 3 presents the bar chart that illustrates the distribution of the examined variables among DHF patients at RSUD Abdul Moeloek Lampung. The bar chart serves to provide a clear comparison of the proportion of patients with normal and abnormal levels of hemoglobin, leukocytes, and hematocrit. By visualizing the data in this manner, it becomes easier to observe patterns and disparities across the three hematological indicators, which subsequently offer an initial descriptive understanding before proceeding to further statistical analysis. This graphical representation not only highlights the dominance of certain categories but also facilitates interpretation regarding the potential role of these variables in influencing the clinical profile of DHF patients.

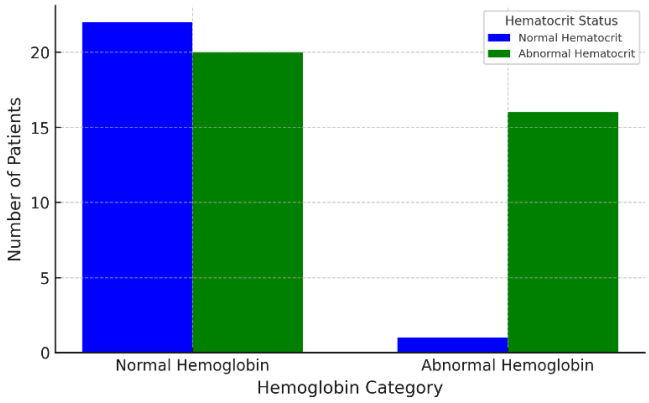

Figure 3. Bar chart of hemoglobin and hematocrit.

Patients with normal hemoglobin levels tended to exhibit slightly higher proportions of normal hematocrit levels compared to abnormal ones, indicating that most patients with normal hemoglobin levels also presented with normal hematocrit levels. Conversely, among patients with abnormal hemoglobin levels, the proportion of abnormal hematocrit levels was significantly higher than that of normal hematocrit levels. These findings suggest a potential association between hemoglobin and hematocrit levels. To further examine this relationship, the chi-square test was conducted, and the results are presented in Table 3.

Table 3. Chi-Square Tests 2

Test	Value	*df*	asymp. Sig. (2-sided)
Pearson Chi-Square	11,000a	1	.001
Likelihood Ratio	13.168	1	.000
Linear-by-Linear Association	10,814	1	.001
N of Valid Cases	59		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 6.63.
b. Computed only for a 2x2 table

The chi-square value was 11,000 with a degree of freedom of 1 and a *p*-value of 0.001 (less than 0.05). Because the *p*-value obtained is less than a significant level of 0.05, then H_0 is rejected. It can be concluded that there is a relationship between hemoglobin and hematocrit. The results of the chi-square test are also in line with those presented in the Fig. 3, showing that there is a relationship between hemoglobin and hematocrit.

4.1.3. Relationship between Leukocytes and Hemoglobin

Similar to the analysis presented in Sections 4.1.1 and 4.1.2, the relationship between leukocyte levels and hemoglobin levels is first illustrated in Fig. 4. This visualization provides a descriptive overview of how variations in leukocyte counts correspond to hemoglobin status among DHF patients, serving as a preliminary step before statistical testing. The bar chart is presented in Fig. 4.

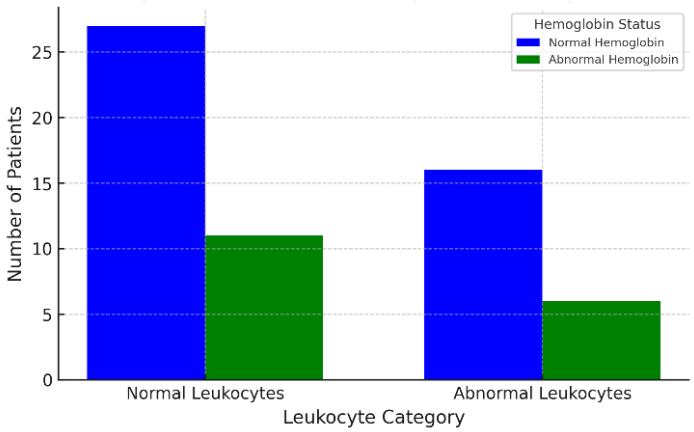

Fig. 4. Bar chart of leukocytes and hemoglobin.

It can be observed that among patients with normal leukocyte levels, the proportion of normal hemoglobin levels was substantially higher than that of abnormal hemoglobin levels, indicating that most patients with normal leukocyte counts also exhibited normal hemoglobin levels. A similar distribution was also found among patients with abnormal leukocyte levels, where normal hemoglobin levels were still more prevalent than abnormal ones. This consistent pattern across both categories suggests that variations in leukocyte levels do not substantially affect hemoglobin status. Therefore, it can be inferred that there is no significant relationship between leukocyte levels and hemoglobin levels.

Table 4. Chi-Square Tests 3

Test	Value	df	asymp. Sig. (2-sided)
Pearson Chi-Square	.041a	1	.840
Likelihood Ratio	.041	1	.840
Linear-by-Linear Association	.040	1	.842
N of Valid Cases	59		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 6.63.
b. Computed only for a 2x2 table

Based on Table 4, the pairwise chi-square test between leukocyte and hemoglobin categories is non-significant ($\chi^2 = 0.041,\ df = 1,\ p = 0.841$), although the interaction terms were evaluated and included or excluded in the final log-linear model based on hierarchical likelihood ratio (ΔG^2) comparisons between nested models rather than pairwise tests alone. Unlike the previous test results, the *p*-value obtained was greater than the significance threshold, indicating that the null hypothesis (H_0) cannot be rejected. This finding is consistent with the descriptive analysis from Fig. 4, which suggested no association between leukocyte levels and hemoglobin levels. Therefore, it can be inferred that there is a high possibility that no interaction exists between leukocyte levels and hemoglobin levels in the best-fitting model.

4.2. Model Analysis

After conducting bivariate tests between the variables, the next step was to perform modeling and select the best-fitting model showing the relationship between the observed variables and the incidence of DHF. Using SPSS software, the modeling process was carried out systematically, and the results are presented in Table 5.

Table 5. Model Selection

No	Model	G^2	*df*	sig
1	$L + Hb + H + Hb * H + L * H + L * Hb + L * Hb * H$	0.000	0	.
2	$L + Hb + H + Hb * H + L * H + L * Hb$	0.688	1	0.407
3	$L + Hb + H + L * Hb * H$	0.000	0	.
4	$L + Hb + H$	20,896	4	0.000
5	$L + Hb + H + L * Hb$	20,855	3	0.000
6	$L + Hb + H + L * Hb + L * H$	14,900	2	0.001
7	$L + Hb + H + L * H + Hb * H$	1,773	2	0.412
8	$L + Hb + H + L * Hb + Hb * H$	7,688	2	0.021
9	$Hb * H + L * Hb + L * H$	0.688	1	0.407
10	$Hb * H + L * Hb + L * H + L * Hb * H$	0.000	0	.
11	$Hb * H + L * Hb + L * H + L$	0.688	1	0.407
12	$Hb * H + L * Hb + L * H + L + Hb$	0.688	1	0.407
13	$Hb * H + L * Hb + L * H + L + H$	0.688	1	0.407
14	$Hb * H + L * Hb + L * H + Hb + H$	0.688	1	0.407

*L=Leukocytes, Hb=Hemoglobin, and H=Hematocrit

It can be seen that 14 combinations of models were studied and compared. Model 1 ($L + Hb + H + Hb * H + L * H + L * Hb + L * Hb * H$), model 2 ($L + Hb + H + L * Hb * H$), and model 3 ($Hb * H + L * Hb + L * H + L * Hb * H$) had G^2= 0, with a significant value of 0. Therefore, this model is not very appropriate to use. The three models appear unsuitable because they are relatively complex, employing the interaction of three variables at once. It can also be inferred that incorporating the interaction of three variables at once will result in the model not being suitable for use. So, that further analysis was carried out without using interactions of three variables at once.

Next, other simpler models were compared without incorporating the interaction of three variables. Model 4 ($L + Hb + H$) was the simplest model because it did not involve interaction between variables. However, this model yielded the largest value, which was 20.896, with a significant value of 0.000 (less than 0.05). This result showed that this model could not be used because its significant value was 0.000, leading to the rejection of the model. Model 5 ($L + Hb + H + L * Hb$) was more complex than model 4, as it incorporated the interaction between leukocytes and hemoglobin. This model also produced a large value, which was, 20.855 with a significance value of 0.000 (less than 0.05). This result indicated that this model could not be used because its significant value was 0.05, leading to the rejection of the model. Model 6 ($L + Hb + H + L * Hb + L * H$) and model 8 ($L + Hb + H + L * Hb + Hb * H$) were more complex than model 5. However, they yielded smaller G^2 value of 14,900 and 7,688, respectively. Then, model 6 had a significant value of 0.00, while model 8 had a significant value of 0.021 (both less than 0.05). Similar to model

4 and model 5, model 6 and model 8 had a significant value of less than 0.05, indicating rejection of H_0 and suggesting that these models are also not suitable.

Interestingly, the most complex models with many interactions were not suitable. Likewise, the simplest models without interaction were found to be unsuitable. Hence, a combination of other models was carried out until a good model was obtained. Furthermore, more complex models than model 4, namely model 2 ($L + Hb + H + Hb * H + L * H + L * Hb$), model 9 ($Hb * H + L * Hb + L * H$), model 11 ($Hb * H + L * Hb + L * H + L$), model 12 ($Hb * H + L * Hb + L * H + L + Hb$), model 13 ($Hb * H + L * Hb + L * H + L + H$), and model 14 ($Hb * H + L * Hb + L * H + Hb + H$) had similar G^2, degrees of freedom, and p-value. The value of G^2 was 0.688, with a degree of freedom of 1, and a p-value of 0.407. These results suggested that these six models were appropriate because they had a p-value that was greater than the 0.05 significant level, so H_0 was not rejected. These results are quite interesting because with several combinations of models obtained the exact same value. Model 2, model 9, model 11, model 12, model 13, and model 14 had an interaction between leukocyte level and hemoglobin level. However, as previously explained, there is no relationship between leukocytes and hemoglobin, which may explain why these six models yielded the same values.

Therefore, the six previous models were compared with a model with no interaction between leukocytes and hemoglobin, namely model 7 ($L + Hb + H + L * H + Hb * H$). Model 7 was not too complex but not too simple. It had G^2 of 1.773, with degrees of freedom 2 and a p-value of 0.412. Model 7 had a p-value that was also greater than 0.05, meaning that H_0 is not rejected and the model was deemed appropriate. If compared to the six previous models, model 7 produced a higher p-value. Therefore, that in this study, model 7 was the selected G^2, which had a fairly small value with the largest p-value. Thus, it can be concluded that model 7 is the most suitable model or the best among other model combinations. Then from the selected model, parameter estimation was carried out. The results are shown in Table 6.

Table 6. Estimate Parameter

Parameter	Estimate	Std. Error	Z	Sig.
Constant	1.386a			
[Leukocytes = 1]	1,099	.385	2.854	.004
[Leukocytes = 2]	0b	.	.	.
[Hemoglobin = 1]	.223	.335	.665	.506
[Hemoglobin = 2]	0b	.	.	.
[Hematocrit = 1]	-1.957	1.086	-1.802	.072
[Hematocrit = 2]	0b	.	.	.
[Leukocytes = 1] * [Hematrocytes = 1]	-1.361	.570	-2.387	.017
[Leukocytes = 1] * [Hematrocytes = 2]	0b	.	.	.
[Leukocytes = 2] * [Hematrocytes = 1]	0b	.	.	.
[Leukocytes = 2] * [Hematocytes = 2]	0b	.	.	.
[Hemoglobin = 1] * [Hematrokit = 1]	2.868	1.076	2.665	.008
[Hemoglobin = 1] * [Hematrokit = 2]	0b	.	.	.
[Hemoglobin = 2] * [Hematrokit = 1]	0b	.	.	.
[Hemoglobin = 2] * [Hematrokit = 2]	0b	.	.	.

The parameter estimates from the final log-linear model are presented in Table 6. Leukocytes in the abnormal category were significantly associated with increased expected cell frequencies compared to the reference category ($\beta = 1.099$, $p = 0.004$; $exp(\beta) = 3.00$, 95% CI: 1.43–6.29). Hemoglobin showed no significant effect ($\beta = 0.223$, $p = 0.506$), while hematocrit showed a marginal negative effect ($\beta = -1.957$, $p = 0.072$). A significant negative interaction between leukocytes and hematocrit ($\beta = -1.361$, $p = 0.017$; $exp(\beta) = 0.26$, 95% CI: 0.09–0.77) suggests that the co-occurrence of abnormal leukocyte and hematocrit levels was associated with lower-than-expected cell frequencies. In contrast, the interaction between hemoglobin and hematocrit was significantly positive ($\beta = 2.868$, $p = 0.008$; $exp(\beta) = 17.6$, 95% CI: 2.1–146.2), indicating a

strong positive association. Reference categories ("0b") correspond to the baseline levels for each variable.

5. Conclusion

Based on the explanation and findings, it can be concluded that leukocyte, hemoglobin, and hematocrit categories are significantly associated with each other within the DHF patient sample, with notable two-factor interactions between leukocyte–hematocrit and hemoglobin–hematocrit. Furthermore, the interaction effects between leukocytes and hematocrit as well as between hemoglobin and hematocrit were also found to significantly contribute to the model, indicating that the combined influence of these variables provides additional explanatory power in understanding the occurrence of DHF cases. However, the interaction between leukocytes and hemoglobin did not demonstrate a significant effect and therefore was excluded from the final model. The best-fitting model obtained from the analysis was expressed as $L + Hb + H + L * H + Hb * H$. This implies that DHF cases in Lampung Province are strongly associated with leukocyte, hematocrit, and hemoglobin levels, both individually and in terms of specific interactions. Nevertheless, it is important to note that leukocyte and hemoglobin levels do not directly influence one another, meaning that a decrease in leukocyte levels does not necessarily correspond to a decrease in hemoglobin levels, and vice versa. These findings provide a more comprehensive understanding of the hematological factors influencing DHF and can serve as a reference for further clinical or epidemiological studies.

Acknowledgment

Thank you to Abdul Moeloek Hospital in Lampung for allowing us to conduct research on dengue cases.

References

[1] O.A. Odetunmibi, A.O. Adejumo, and T.A. Anake, "Log-linear modelling of effect of age and gender on the spread of hepatitis B virus infection in Lagos State, Nigeria," *Open Access Maced. J. Med. Sci.*, vol. 7, no. 13, pp. 2204–2207, 2019, doi: 10.3889/oamjms.2019.573.

[2] M. Teinsuwan, S. Lertprapai, J. Sirichaisinthop, and A. Lawmepol, "Application of log-linear models to malaria patients in Thailand," *Stat. Med.*, vol. 19, no. 14, pp. 1931–1945, 2000, doi: 10.1002/1097-0258(20000730)19:14<1931::aid-sim504>3.0.co;2-t.

[3] B.J. Swihart, B.S. Caffo, C.M. Crainiceanu, and N.M. Punjabi, "Mixed effect Poisson log-linear models for clinical and epidemiological sleep hypnogram data," *Stat. Med.*, vol. 31, no. 9, pp. 855–870, 2012, doi: 10.1002/sim.4457.

[4] J.W. Millard, F.A. Alvarez-Nunez, and S.H. Yalkowsky, "Solubilization by cosolvents establishing useful constants for the log-linear model," *Int. J. Pharm.*, vol. 245, no. 1–2, pp. 153–166, 2002, doi: 10.1016/s0378-5173(02)00334-4.

[5] S.F. Sihotang and Z. Zuhri, "Analisis model log linear tiga dimensi untuk data kualitatif dengan metode forward," *J. Math. Edu. Sci.*, vol. 6, no. 1, pp. 62–69, 2020, doi: 10.30743/mes.v6i1.3137.

[6] T.A. Nurman, "Analisis data kategori dengan log linier menggunakan prinsip hirarki (studi kasus jumlah kecelakaan lalu lintas di Kota Makassar tahun 2011)," *Teknosains Media Inf. Sains Teknol.*, vol. 7, no. 1, pp. 99–110, 2013, doi: 10.24252/TEKNOSAINS.V7I1.77.

[7] I. Kamil, M. Susilawati, and I.P.E.N. Kencana, "Model log-linear faktor-faktor yang mempengaruhi hipertensi (studi kasus: RSUD Abdoe Rahem Situbondo)," *E-J. Mat.*, vol. 1, no. 1, pp. 84–88, 2012, doi: 10.24843/MTK.2012.v01.i01.p015.

[8] J.K. Vermunt, "LEM: Log-Linear Modelling, User's manual," in *Encyclopedia of Behavioral Statistics*, B. Everit and D.C. Howell, Eds. Chichester, United Kingdom: John Wiley & Sons Inc, 2005, pp. 1082-1093.

[9] M. Bucca and D.R. Urbina, "Lasso regularization for selection of log-linear models: An application to educational assortative mating," *Sociol. Methods Res.*, vol. 50, no. 4, pp. 1763-1800, 2019, doi: 10.1177/0049124119826154.

[10] A. Hakami et al., "Hematological parameters in recent and past dengue infections in Jazan Province, Saudi Arabia," *Saudi Med. J.*, vol. 46, no. 5, pp. 515–521, May 2025, doi: 10.15537/smj.2025.46.5.20240879.

[11] S. Anwar, R. Islam, Y. Araf, J.T. Mourosi, and M.J. Hosen, "Clinical profiles and hematologic patterns in Bangladesh's 2019 dengue outbreak," *The Microbe*, vol. 7, Jun. 2025, Art. no 10032, doi: 10.1016/j.microb.2025.100325.

[12] Mst.S. Akhtar, Md.M. Rahman, and Mst.S. Akhter, "Observation of hematocrit level among dengue fever patients in a Tertiary Care Hospital," *Sch. Int. J. Anat. Physiol.*, vol. 8, no. 3, pp. 67–72, May/Jun. 2025, doi: 10.36348/sijap.2025.v08i03.001.

[13] S. Bhatt et al., "The global distribution and burden of dengue," *Nature*, no. 496, pp. 504–507, 2013, doi: 10.1038/nature12060.

[14] O.J. Brady et al., "Refining the global spatial limits of dengue virus transmission by evidence-based consensus," *PLoS Negl. Trop. Dis.*, vol. 6, no. 8, 2012, Art. no. e1760, doi: 10.1371/journal.pntd.0001760.

[15] A.S. Mufidah and Purhadi, "Analisis survival pada pasien demam berdarah dengue (DBD) di RSU Haji Surabaya menggunakan model regresi Weibull," *J. Sains Seni ITS*, vol. 5, no. 2, pp. D301–D306, 2016, doi: 10.12962/j23373520.v5i2.16633.

[16] WHO, "Dengue and Severe Dengue," 2025. [Online]. Available: https://www.who.int/en/news-room/fact-sheets/detail/dengue-and-severe-dengue

[17] M. Irfan, M. Usman, Warsono, D. Kurniasari, and Widiarti, "Survival analysis using cox proportional hazard regression approach in dengue hemorrhagic fever (DHF) case in Abdul Moeloek Hospital Bandar Lampung 2019," in *Proc. 3rd Int. Conf. Appl. Sci. Math. Inform*., in Journal of Physics: Conference Series, vol. 1751, 2020, doi: 10.1088/1742-6596/1751/1/012011.

[18] A. Agresti, *An Introduction to Categorical Data Analysis*. New York, USA: John Wiley & Sons, 2007.

[19] Maryana, "Model log linier yang terbaik untuk analisis data kualitatif pada tabel kontingensi tiga arah," *Malikussaleh Indus. Eng. J.*, vol. 2, no. 2, pp. 32–37, 2013, doi: 10.53912/iejm.v2i2.77.

[20] S.B. McKenzie, J.L. Williams, and K. Landis-Piwowar. *Clinical Laboratory Hematology*. Boston, MA, USA: Pearson, 2020.

[21] M. Riaz et al., "Evaluation of clinical and laboratory characteristics of dengue viral infection and risk factors of dengue hemorrhagic fever: A multi-center retrospective analysis," *BMC Infect. Dis.*, vol. 24, no. 1, May 2024, Art. no. 15, doi: 10.1186/s12879-024-09384-z.

[22] E. Mulyani, S. Nugroho, and F. Faisal, "Model log linier beberapa kasus kriminologi yang terjadi di wilayah Polres Bengkulu pada tahun 2004/2005," *J. Alumni Jur. Mat. Staf Pengajar Jur. Mat. FMIPA Univ. Bengkulu*, 2006.