ENTHUSIASTIC
INTERNATIONAL JOURNAL OF APPLIED STATISTICS AND DATA SCIENCE
Volume 4, Issue 2, October 2024, pp. 119-131

Statistical Perspective of Dengue Hemorrhagic Fever in West Java: Insights from Two-Way RE Model

Ghiffari Ahnaf Danarwindu [a,1,*], Muhammad Ghani Fadhlurrahman [b,2]
[a] Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Islam Indonesia, Sleman, Indonesia
[b] Alumni of the Master of Mathematics Study Program, Universitas Gadjah Mada, Sleman, Indonesia
[1] ghiffari.ahnaf@uii.ac.id*; [2] mghanif1109@gmail.com
* Corresponding author

ARTICLE INFO

Keywords
panel data
dengue hemorrhagic fever (dhf)
two-way random effect model
biostatistics
regression

ABSTRACT

The Indonesian Ministry of Health has reported an alarming increase in Dengue Hemorrhagic Fever (DHF) cases, particularly in West Java Province. Given this trend, collaborative research and surveillance efforts are crucial to understanding and managing DHF cases in Indonesia. The panel data regression model in dengue fever cases will provide new insights into modeling. This research aimed to identify the most appropriate random effects model for estimating a dataset with four different variables. This study involved panel data variables on the effect of population density, percentage of poor people, percentage of households with access to clean water, and proper sanitation on DHF cases in West Java Province. This method emphasized selecting the best model from one-way and two-way Random Effects (RE) models and identifying what factors influenced the increase of DHF cases in West Java province. The best model obtained was a two-way RE Model with three significant variables. Based on the selected variables in the model, West Java Province needs to pay attention to the distribution of housing and economic activity in each district because population density is a crucial concern for the local government.

1. Introduction

Dengue Hemorrhagic Fever (DHF) is a tropical disease caused by the dengue virus (DENV). This virus is primarily contracted when mosquitoes, particularly Aedes aegypti, nibble on humans. These mosquitoes transmit the virus via their saliva during the biting process, allowing healthy individuals to become exposed. The Ministry of Health of the Republic of Indonesia has reported an alarming increase in dengue fever cases, particularly on the island of Java. In West Java province alone, 39,623 confirmed cases were confirmed, making it the highest incidence in the country [1]. Given this trend, collaborative research and surveillance efforts are crucial to understanding and managing DHF cases in Indonesia.

The objectives of data analysis are to present information to make it easier to collect and analyze data [2]. In epidemiology, especially in the study of infectious diseases, providing appropriate information is useful for speeding up treatment. DHF cases in West Java

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

Province require further data analysis, especially regarding the spread of the disease, to find out what factors affect the increase of DHF in regions and cities in West Java. Studies related to distribution patterns and factors influencing DHF disease in cities in West Java have an interesting perspective to research. Much research in the statistical and mathematical field has been conducted on modeling and analyzing data about dengue fever, especially cases in Indonesia. From a health perspective, many research and clinical trials have been carried out to find parameters for the spread of dengue fever [3], [4].

Mathematical and statistical approaches have also been discussed in several studies. One of the studies about modeling focused on the analysis of dengue fever disease models by looking at the parameters of infected individuals, the transition between susceptible and infected individuals, and then recovered individuals [4]. Meanwhile, others focused on quantitative methods of different views on seeing dengue fever, and the preventive action for vulnerable groups was calculated using the chi-square test and logistic regression [5]. However, from a statistical modeling perspective, particularly in panel data regression models, dengue fever is rarely discussed.

Panel data models in several situations provide better results than cross-sectional data or time series data, including more information, more degrees of freedom, and less linear multiplicity between variables. Problems related to heterogeneity in data from cross-sectional objects or time-series objects will be resolved in panel data models with certain variants [6], [7]. In simple terms, panel data models are divided into three types: pooled OLS models, often referred to as common effect (CE) models; Fixed Effect (FE) models; and Random Effect (RE) models, which are specifically further varied into six different types [8]. The FE and RE models are formed from a combination of errors from the cross-section and time series, commonly called the Error Component Model (ECM) [9]. In practice, the FE model centered on the intercept, while the RE model centered on the error [10]. In this study, researchers wanted to examine RE models in two directions. It can be said that both individuals and time affect the assessment of the selected variables significantly. Based on the facts and researcher opinions, this research provided a different perspective to identify a two-way random effect model for estimating a dataset with four different variables.

The research project analyzed data on individuals and time spanning the years 2018 to 2023. The dataset consisted of panel data, which combined cross-sectional and time series information [11]. Panel data regression techniques were employed to track the data across multiple periods. The key benefit of using panel data regression is its ability to account for variations between different cross-sectional units, providing more comprehensive insights compared to straightforward time-series analysis [12]. Extensive research has utilized panel data analysis, investigating factors that impact poverty rates within specific regions [13], [14]. However, there has been limited exploration of panel data analysis concerning the determinants of specific diseases. One of the studies related to panel data analysis for the health sector was recently carried out on DHF cases in Bogor district [15]. The research gap that is intended to be shown with the previous model is by expanding the scope of the research area, it is expected that the panel data model found will be different and produce different interpretations of the predetermined variables. The difference in variables with previous research is also a determining factor in the research gap because different perspectives will be found regarding the results of the model and interpretation of the panel data regression analysis to be studied. This became the basis for research related to factors influencing DHF disease in West Java province using panel data. We believe that the panel data regression model in dengue fever cases will provide new insights in modeling and the

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

field of panel data in the case of dengue fever. Especially from the government side, the statistical approach to DHF data will be one of the benchmarks for mitigation planning for clean water supply, urbanization, and population distribution. These indicators are selected based on the variables used in panel data analysis for DHF cases in West Java province.

2. Method

2.1. Data Variables

The variables used in modeling data panel regression in DHF disease cases divided into response and predictor variables. The response variable was the number of dengue fever cases in the districts and cities of West Java province in 2018–2023. Meanwhile, the predictor variables were a possibility that can cause the number of DHF cases in West Java province from 2018-2023. Details of the variables obtained are presented in Table 1.

Table 1. Data variables of data panel regression models

Variable	Description	Unit	Measure
Y	The number of DHF cases in West Java Province	Percent	Ratio
X_1	Population Density of each Regency and City in West Java Province	Kilo meters square	Ratio
X_2	Percentage of Poor Population by Regency/City	Percent	Ratio
X_3	Percentage of Households with Access to Safe Water	Percent	Ratio
X_4	Percentage of Households with Access to Adequate Sanitation by Regency/City	Percent	Ratio

The span between 2018 and 2023 was chosen because it was the closest year that research was taken. In addition, one condition that a two-way remodel can be processed is that the number of times or years chosen must not exceed the number of independent variables. Then, all independent variables were chosen because there are main responsible factors for the Increase DHF cases in some area. Two major factors are drastic population growth and substandard housing, crowding, and deterioration in water, sewer, and waste management associated with unplanned urbanization [16]. Those factors led the appropriate condition for aedes aegypti mosquitoes transmitted the disiases in some areas.

2.2. Panel Data Regression Model

The panel data regression model is composed of a panel data structure that forms from observations of a similar cross-section unit over time [8]. The common panel data regression model equation is [17]:

$$y_{it} = \alpha + \boldsymbol{X}_{it}\boldsymbol{\beta} + \varepsilon_{it} \quad (1)$$

with y_{it} is variable response of i number of individuals and t number of times, α is coefficient of intercept, $\boldsymbol{X}$ is k number of predictor variable of i number of individuals and t number of times, $\boldsymbol{\beta}$ is slope coefficient with size $k \times 1$, and ε_{it} is regression residuals of i number of individuals and t number of times.

2.3. Multicollinearity Detection

The multicollinearity can be detected with the Variance Inflation Factor (VIF) test, so the level of collinearity or correlation between predictor variables can be seen. Hence, variables are good enough to be processed because each variable is not similar. The VIF test as follow [18]:

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

where R_j^2 is the contribution of independent variables that generated from the regression of variable x_j with other independent variables and $j = 1,2, \dots, k$ is the independent variables that contained in the study. If the output test is smaller than 10, it can be stated that there is no multicollinearity.

$$VIF = \frac{1}{R_j^2} \quad (2)$$

2.4. Estimation of Panel Data Regression Model

There are three approaches to estimating panel data regression models, namely the CE model, FE model, and RE model [8].

a. CE Model

CE model has a similar estimate to that of common linear regression models. The main principle of the CE model is to combine the model data without considering individuals and time. In addition, the intercept alpha (α) is considered exactly the same or constant in each individual and time. In general, the CE model regression model is [9], [19]:

$$y_{it} = \beta_0 + \beta_1 x_{1it} + \beta_2 x_{2it} + \cdots + \beta_k x_{kit} + \varepsilon_{it} \quad (3)$$

b. FE Model (One-Way and Two-Way)

FE Model is a panel regression in which estimation can be differentiated based on individuals and time. The parameter estimation worn in the fixed effect approach is Least Square Dummy Variable (LSDV), a method used in estimating linear regression parameters using OLS in models that involve dummy variables as predictor variables for different intercepts for each individual and time. Here are some types of FE models [9], [19]:

- FE model constant slope coefficient with different intercept coefficients for each individual.

  The FE model on individuals has a constant coefficient slope value, and the intercept coefficient will be different for each individual, assuming the time effect is not included. Each individual has the assumption that the time effect is ignored. This condition states that even the intercept can change for individuals which are loaded, each intercept does not change over time or time does not change. The regression model for the individual FE model is:

  $$y_{it} = D_i \alpha_i + \beta_1 x_{1it} + \beta_2 x_{2it} + \cdots + \beta_k x_{kit} + \varepsilon_{it} \quad (4)$$

  with D_i is dummies for the *i* number of individuals and α_i is the slope of the *i* number of individuals.

- FE model constant slope coefficient with different intercept coefficients at each time. The FE model over time has an unchanged slope coefficient value and the intercept coefficient not same over time assuming individual effects are not included. The value of the intercept coefficient can be different over time because there are dummy variables categorized by time. The regression model for the FE model of time is:

  $$y_{it} = D_t \lambda_t + \beta_1 x_{1it} + \beta_2 x_{2it} + \cdots + \beta_k x_{kit} + \varepsilon_{it} \quad (5)$$

  with D_t is dummy for the *t* number of times and λ_t is the slope of the *t* number of times.

- FEM constant slope coefficient with different intercept coefficients for each individual and time.

  The last one is a FE model that incorporates both individual and time effects. This model has a steady slope coefficient value, and the intercept coefficient can be

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

different across individuals and time. The value of the intercept coefficient can differ between individuals and time because there are dummy variables categorized by individual and time. The regression model for the individual and time FE model is:

$$y_{it} = D_i\alpha_i + D_t\lambda_t + \beta_1 x_{1it} + \beta_2 x_{2it} + \cdots + \beta_k x_{kit} + \varepsilon_{it} \quad (6)$$

c. RE Model

There is a chance that individual effects are not observed in the FE model, so the data may be modeled as the RE model. This model may only appear to be a model that adds cross-sectional units to the study and not out-of-sample variables [10]. The benefit of using this RE model is that it eliminates the need to look for correlation and heteroscedasticity. The equation RE model is:

$$y_{it} = \alpha + \beta_1 x_{1it} + \beta_2 x_{2it} + \cdots + \beta_k x_{kit} + u_{it} \quad (7)$$

with $u_{it} = v_i + e_t + \varepsilon_{it}$, where v_i affects all observations for cross-section unit i, e_t affects all observations for time t, and ε_{it} affects only observation it.

2.5. Selection of The Best Model

To understand which regression model to choose for analysis, it is good to test the model specification. The following are some of the tests carried out [10].

2.5.1. Chow Test

The Chow test is the most common choice to comparing the CE model and FE model. Here are the hypotheses and tests [8], [12]:

$H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0$ (CE model is the best model)

H_1 : at least there is one $\beta_j \neq 0, j = 1,2, \dots, k$ (FE model is the best model)

If $F \geq F_{\alpha,df_1,df_2}$ with $df_1 = n - 1$ and $df_2 = nt - n - k$, then H_0 is rejected. The statistic test as follow as

$$F = \frac{(R^2_{LSDV} - R^2_{Pooled})/(n-1)}{(1-R^2_{LSDV})/(nt-n-k)} \quad (8)$$

where R^2_{LSDV} is the FEM R^2 value, R^2_{Pooled} is the CEM R^2 value, n is number of inidivual, t is number of times, and k is number of predictor variables.

2.5.2. Hausman Test

The Hausman test allows researchers to choose the top model from FE model and RE model. Here are the hypotheses and tests [8], [12]:

$H_0 : corr(x_{it}, \varepsilon_{it}) = 0$ (REM is the best model)
$H_1 : corr(x_{it}, \varepsilon_{it}) \neq 0$ (FEM is the best model)
If $W > \chi^2_{\alpha,df}$ with $df = k - 1$, then H_0 is rejected. The statistic test as follow as

$$W = (\boldsymbol{b} - \widehat{\boldsymbol{\beta}})' \left(var(\boldsymbol{b}) - var(\widehat{\boldsymbol{\beta}})\right)^{-1} (\boldsymbol{b} - \widehat{\boldsymbol{\beta}}) \quad (9)$$

where $\boldsymbol{b}$ is matrix of beta from fixed effect model and $\widehat{\boldsymbol{\beta}}$ is matrix of beta from random effect model.

2.5.3. Lagrange Multiplier Test

The Lagrange multiplier test is analyzed to assist in the determination of whether the CE model or the RE model is more appropriate. Here are the hypotheses and tests [8], [12]:

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

$H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2 = \sigma^2$ (CE model is the best model)
$H_1 : \sigma_i^2 \neq \sigma^2$ (RE model is the best model)
If $LM > \chi^2_{\alpha,df}$ with $df = n - 1$, then H_0 is rejected. The statistic test as follow as

$$LM = \frac{nt}{2(t-1)}\left(\frac{\sum_{i=1}^{n}\left(\sum_{t=1}^{t}\varepsilon_{it}\right)^2}{\sum_{i=1}^{n}\sum_{t=1}^{t}\varepsilon_{it}^2} - 1\right)^2 \quad (10)$$

2.6. Significance Parameter Test

Parameter significance testing is done because it helps to see the level of effect the predictor variables have on the response variable. Parameter significance testing is done simultaneously and partially. The explanation of each test is as follows.

2.6.1. Simultaneous Test

This test is performed simultaneously to ensure that the selected model previously has minimum one significance variable. The hypothesis stated as follows [20]:

$H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0$ (all predictor variables almost sure the effect not significance on response variable)

H_1 : There is at least one $\beta_j \neq 0, j = 1,2,\ldots,k$ (there is possibility the model contained one from j number of predictor variables that has an effect very significant on the response variable)

If the significance level alpha (α) is given, then H_0 is rejected if $F \geq F_{\alpha,df_1,df_2}$ with $df_1 = k$ and $df_2 = nt - k - 1$. The statistic test is as follows [20]:

$$F = \frac{MSR/k}{MSE/nt-k-1} \quad (11)$$

2.6.2. Partial Test

This test is performed to implies whether there is a partial effect of the predictor variable, namely β_j for each $j = 0,1,2,\ldots,k$ on the response variable. The hypothesis stated as follows [20]:

$H_0 : \beta_j = 0$

(The j number of predictor variables almost sure the effect not significant on the response variable)

H_1 : There is at least one $\beta_j \neq 0, j = 1,2,\ldots,k$

(The j number of predictor variables has a variable that give significant effect on the response variable)

If given a significance level of α, then H_0 is rejected if $|t_j| \geq t_{\alpha/2,df}$ with $df = n - k - 1$. The statistic test is as follows [20]:

$$t_j = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \quad (12)$$

with $se(\hat{\beta}_j)$ is the standard error obtained from the root variance of the regression parameter estimates.

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

2.7. One-Way RE Model

RE model estimates panel data where disturbance terms have the opportunity to be interconnected over time and individual. The advantage of using this model is its ability to eliminate heteroscedasticity. This model can also be called Error Component Model (ECM) or the Generalized Least Squares (GLS) technique). The general RE model can be seen in (7), but with $u_{it} = v_i + e_t + \varepsilon_{it}$ have two characteristics. First, if the RE model have just individual effect (one-way individual RE model), then $u_{it} = v_i + \varepsilon_{it}$. If the RE model have just time effect (one-way time RE model), then $u_{it} = e_t + \varepsilon_{it}$. So, there are two disturbance terms either in the one-way individual or time RE model.

RE model have several assumptions that errors are not correlated and there is no autocorrelation between cross-section and time series units. This is because GLS method used in this model that caused required assumptions in OLS method (homoscedasticity and non-autocorrelation) are not met [22]. These results open the boundaries for that assumptions or can be concluded either the assumptions are met or not, we still can have the good but not perfect panel data model. Hence, GLS method gave opportunity panel data model not always has to be CE model dan FE model with OLS method.

2.8. Two-Way RE Model

In the panel regression estimation section, it has been explained about the RE model with the error in equation (7) has three error units, each of which carries the effect of the data. The two-way error component model is the regression model in equation (1) but with a two-way error component disturbance [21]:

$$u_{it} = v_i + e_t + \varepsilon_{it} \quad (13)$$

where v_i is the symbol for the unobservable individual effect, λ_t is the symbol for the unobservable time effect and ε_{it} is the symbol for the remaining stochastic disturbance term. Furthermore, if all $v_i \sim IID(0, \sigma_v^2)$, $\lambda_t \sim IID\left(0, \sigma_\lambda^2\right)$, and $\varepsilon_{it} \sim IID(0, \sigma_\varepsilon^2)$ are independent, then this is the definition of the two-way RE model [17]. Moreover, $\boldsymbol{X}_{it}$ is independent of v_i, λ_t, and ε_{it} for all i and t. The conclusion in this case is related to the large population from which the sample was randomly obtained.

Therefore, the two-way RE model can be simplify as the extended of RE model by included time-specific effects. Like the cross-section effects, these time effects are also unobserved and assumed to be uncorrelated with the independent variables. It can also be concluded that both effects are important to look forward to being explained with significant variables to response variable.

3. Results and Discussion

3.1. Descriptive Statistical Analysis

Descriptive statistical analysis is included to make it easier to see the general state of the data such as average, standard deviation, minimum data, maximum data and the amount of data. The results can be seen in Table 2.

Table 2. Descriptive Statistics of the Number of DHF in West Java 2018-2023

Variable	Mean	Median	Minimum	Maximum	Std. Dev
The Number of DHF cases	56.33	40.13	2.44	310	52.76502
Population Density	3919.2	1439	379.9	15643	4643.754
Percentage of Poor Population	8.262	8.265	2.07	13,13	2.750715

Variable	Mean	Median	Minimum	Maximum	Std. Dev
Percentage of Households with Access to Safe Water	92.83	95.59	69.2	100	7.033074
Percentage of Households with Access to Adequate Sanitation	72.18	75.93	34.93	98.52	16.37187

3.2. Multicollinearity Detection

VIF score was used to determine whether there is multicollinearity between predictors with a safe limit for the VIF value of 10. Table 3 shows VIF values from the four independent variables used.

Table 3. The Value of Variance Inflating Factor

Variable	x_1	x_2	x_3	x_4
VIF	1.845341	1.711174	1.215605	1.109468

Based on the table, the VIF value shows that there is no multicollinearity between the four independent variables.

3.3. Estimation of Random Effect Model Parameter

This research focused to seek the two-way RE model, hence, the estimation of CE and FE models were excluded. This can be clarified from the result in the model selection. The RE model estimates panel data regression models using GLS, assuming that variables are uncorrelated and have heteroscedasticity in the variance. The following is the REM model of DHF panel data in West Java Province for 2018–2023.

3.3.1. One-Way Individual RE Model

Panel data regression model of DHF in West Java Province in 2018–2023 using the RE model with individual effect is

$$y_{it} = -96.3516860 + 0.0055883x_{1it} + 4.2592682x_{2it} + 1.7343533x_{3it} - 0.9061554x_{4it} + u_{it} \quad (14)$$

The RE model with individual effect had a score of 0.18306 which stated as coefficient of determination (R^2) or it concluded that the model could explain the variation of DHF in West Java Province in 2018–2023 by 18.306%.

3.3.2. One-Way with Time RE Model

Panel data regression model of DHF in West Java Province in 2018–2023 using the RE model with time effect is

$$y_{it} = -80.8461424 + 0.0048049x_{1it} + 2.5261502x_{2it} + 1.9726968x_{3it} - 1.1865642x_{4it} + u_{it} \quad (15)$$

The RE model with time effect has a score of (R^2) which was 0.36998 or the model could explain the variation of DHF in West Java Province in 2018–2023 by 36.998%.

3.3.3. Two-Way RE Model

Panel data regression model of DHF in West Java Province in 2018-2023 with the RE model with individual and time effect is

$$y_{it} = 17.3204042 + 0.0058081x_{1it} + 3.2575410x_{2it} + 1.0191682x_{3it} - 0.9785983x_{4it} + u_{it} + 1.9726968x_{3it} - 1.1865642x_{4it} + u_{it} \quad (16)$$

The RE model with individual and time effect had a score of (R^2) which was 0.17583 or the model could explain the variation of DHF in West Java Province in 2018-2023 by 17.583%.

3.4. Selection of The Best Model

The appropriate RE model that suitable for the data was selected, however, to confirm that the two-way model is optimal, all tests were conducted.

3.4.1. Chow Test

The Chow test was used to determine whether the CE model was better than the FE model. Table 4 shows the results of the Chow test with the RStudio application.

Table 4. Chow Test Result

Description of Tested Variables	Test Result Value
The Value of F	3.1957
Degree of Freedom 1	26
Degree of Freedom 2	131
P-Value	6.852×10^{-6}
The Value of F Table	1.580372

Based in the results, the test statistic value of $F = 3.1957$ was greater than $F_{(0.05;26;131)} = 1.580372$ and the p-value was smaller than 5% significance or 6.852×10^{-6}, so H_0 was rejected. This result suggested that FE model was more appropriate than the CE model.

a. Hausman Test

The Hausman was used to determine whether the FE model was better than the RE model. Table 5 presents the results of the Hausman test with the RStudio application.

Table 5. Hausman Test Result

Description of Tested Variables	Test Result Value
The Value of Chi-square	5.6084
Degree of Freedom	4
P-Value	0.2304
The Value of Chi-square Table	9.487729

Based on the Hausman test results, the Wald test statistic value = 5.6084 was smaller than $\chi^2_{(0.05;4)} = 9.487729$ and the p-value was greater than 5% significance or 0.2304, so H_0 failed to be rejected. This result suggested that the RE model was more suitable than the FE model.

b. Lagrange Multiplier Test

The Lagrange Multiplier was used to determine whether the CE model is better than the RE model. Table 6 shows the Lagrange multiplier test with the RStudio application.

Table 6. Lagrange Multiplier Test 1st Result

Description of Tested Variables	Test Result Value
The Value of Chi-square	64.239
Degree of Freedom	2
P-Value	1.124×10^{-14}
The Value of Chi-square Table	5.991465

Based on the results, the Wald test statistic value was smaller than the chi-square table and the p-value was greater than 5% significance, so it can be concluded that H_0 failed to be rejected. This result suggested that the RE model was more suitable than the CE model. In addition, t the RE model may exhibit either individual or temporal effects, or both. Therefore,

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

Lagrange Multiplier test was done for each effect. Table 7 and Table 8 presents the second and third results of the Lagrange multiplier test, respectively.

Table 7. Lagrange Multiplier Test (Second Results)

Description of One-Way (Individual)	Test Result Value
The Value of Chi-square	22.031
Degree of Freedom	1
P-Value	2.683×10^{-06}
The Value of Chi-square Table	3.841459

Table 8. Lagrange Multiplier Test (Third Results)

Description of One-Way (Time)	Test Result Value
The Value of Chi-square	42.208
Degree of Freedom	1
P-Value	8.204×10^{-11}
The Value of Chi-square Table	3.841459

The results shown that both chi-square value and p-value demonstrated that RE model had individual and time effects. Therefore, two-way RE was the best model for the data.

3.5. Analysis of The Two-Way RE Model

After best model was declared, the parameters needed simultaneous and partial confirmation of their significance. Table 9 presents the analysis results with RStudio.

Table 9. Simultaneous Test Result

Model	Chi-square	P-Value
Two-Way RE Model	33.4943	9.4609×10^{-7}

The chi-square value was 33.4943 or greater than the chi-square table with an independent degree of 4, which was 9.487729. This suggested that independent variables had a significant effect on dengue hemorrhagic fever rates. Next, the analysis conducted employed the partial parameter significance, as shown in Table 10.

Table 10. Partial Test Result

Variable	P-Value
Population Density	0.0001982
Percentage of Poor Population	0.2037385
Percentage of Households with Access to Safe Water	0.0803282
Percentage of Households with Access to Adequate Sanitation	0.0019509

Based on the partial test results, only three variables were significant to the DHF rate, namely population density, percentage of households with access to safe water, and percentage of households with access to adequate sanitation. This was based on the p-value of each variable which was smaller than 1% significance, namely 0.0001982, 0.0803282, and 0.0019509.

3.6. Analysis of the Best Model with Significance Variable

The significance parameter test yielded the desired results. The F-test conclusion showed that the model had at least one significance variable. On the other hand, the t-test showed that there were three significant variables. From here, the model was retested for parameters with variables that were already significant. This test would ascertain if the significant variable would remain or undergo a significance decrease. The results are presented in Table 11 and Table 12.

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

Table 11. Simultaneous Test Result with Significance Variable

Model	Chi-square	P-Value
Two-Way RE	32.3464	$4{,}4237 \times 10^{-7}$

Table 12. Partial Test Result with Significance Variable

Variable	P-Value
Population Density	0.0001635
Percentage of Households with Access to Safe Water	0.0501957
Percentage of Households with Access to Adequate Sanitation	0.0007623

From Table 11 and Table 12, both the model and the significant variables are well matched. Because of that, the two ways RE model DHF rates in West Java for 2018-2023 are

$$y_{it} = 8.4529681 + 0.0045335x_{1it} + 1.1367811x_{2it} - 1.0448278x_{3it} + u_{it} \quad (17)$$

3.7. Discussion

The panel data model analysis began with checking all variables' collinearity with the VIF value. Based on Table 3, all four variables have a VIF value under 2, or there is no chance each variable has the same value if we proceed. Next, the possible models that can be used as the best model, specifically in RE models, were estimated; thus, three models were used: one-way individual, one-way time, and two-way re-model. The test results of the best models are presented in Table 4 until Table 8. From the results, it can be concluded that the most suitable model was the RE model. In addition, all p-values were less than 0.05; hence, the best model was the two-way RE model. The final step was determining the significance of the variables. The last step is seeking the significance variables; with the results from simultaneous and partial tests, only three variables could be concluded in the two-way re-model.

Based on the technical analysis, there was a two-way RE model with three significant variables. Even though the two-way model was suitable, the R-squared value was only 17%, less than one-way individual and one-way time. From a statistical perspective, this model is not good because the three significant variables from collected data over six years cannot solve all the DHF cases in West Java. In other words, more variables must be collected to close the 83% gap. In addition, besides R-squared, MAPE, MAD, and MSD tests were employed to see how good or bad the model in Table 13 is, and the results were similar to those of the R-squared.

This research limitation is that choosing other variables that suit the model is not easy. When some variables are collected, the year must be greater than all variables. Thus, not only variables are needed even after finding data for six years, obtaining complete data is not easy. Even after six years, obtaining complete data is not easy. Even though this is the researcher's concern, the researcher believes these three variables can serve as initial steps to reduce DHF cases in West Java.

Table 13. MAPE, MAD, MSD Test

Test	MAPE	MAD	MSD
One-Way Individual	136.2191	30.36644	1853.407
One-Way Time	142.6084	30.48514	1838.291
Two-Way	139.8960	30.84037	1884.204

4. Conclusion

The perspective analysis obtained from this study illustrates that panel data regression does not always conclude with a one-way FE model, two-way FE model, or one-way remodel. Obtaining a two-way RE model provides a new perspective on looking at the data in more detail individually and over time. This research shows that the dengue fever rate in West Java has been very high for the last 5 (five) years, in terms of the province being one of the five provinces with the largest rate while the district/city is in the first rank.

Provincial and district governments are certainly aware of this. Still, from this research, there are three main points of concern: the density of the population in an area, the access people must clean water, and the need for proper sanitation in every home. The high DHF rate over the past five years has shown that the population is overcrowded because the birth rate grows yearly while the land needed for housing cannot increase. Furthermore, there is a lack of education from various parties about keeping water clean and the importance of proper sanitation because West Java province is a more humid and rainier region, so mosquitoes prefer this situation. With these results, researchers hope that the government can have short-term and long-term programs by paying attention to the aspects found by researchers.

A few variables still limit this study, so the determination of the two-way RE model can still be expanded. Recommendations for further research are indicated to expand the variables affecting the DHF spread in West Java province. Therefore, more accurate and interpretative results can be found based on the perspective of other variables. This research is also expected to be a reference for similar research in other provinces in Indonesia, more broadly in areas with subtropical climates and similar cases.

References

[1] Direktorat Jenderal Pencegahan dan Pengendalian Penyakit, "Info Kasus DBD sampai minggu ke 27 Tahun 2024." Accessed: May 24, 2024. [Online]. Available: https://p2p.kemkes.go.id/update-data-dengue/#_h

[2] J. Jajang, B. Pratikno, and M. Mashuri, "Modeling Dengue Fever by using Conditional Autoregressive Bessag-York-Mollie," *Indonesian Journal of Statistics and Its Applications*, Vol. 6, No. 1, pp. 101–113, 2022, doi: 10.29244/ijsa.v6i1p101-113.

[3] Y.I. Jayadi *et al*, Comparative Study of Food Services Between State and Private Islamic Hospitals in Makassar," *Journal of Health Science and Prevention*, Vol. 7, No. 1, pp. 59–73, 2023, doi: 10.29080/jhsp.v7i1.898.

[4] H. Nishiura, "Mathematical and Statistical Analyses of the Spread of Dengue," *Dengue Bulletin*, Vol. 30, pp. 51–67, 2006.

[5] A. Abbasi, "Dengue Fever: A Statistical Analysis Regarding Awareness about Dengue among University Students in Azad Kashmir," *Journal of Healthcare Communications*, Vol. 2, No. 1, pp. 1–8, 2017, doi: 10.4172/2472-1654.100041.

[6] J.M. Wooldridge, *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA, USA: MIT Press, 2010.

[7] R.A. Yaffee, "Primer for Panel Data Analysis." Accessed: October 8, 2024. [Online]. Available: https://web.pdx.edu/~crkl/ec510/pda_yaffee.pdf

[8] D.N. Gujarati and D.C. Porter, *Basic Econometrics*, 5th ed. New York, NY, USA: McGraw-Hill Inc, 2008.

[9] R. Davidson and J.G. MacKinnon, *Estimation and Inference in Econometrics*. Oxford, England: Oxford University Press, 2021.

[10] D. Asteriou and S.G. Hall, *Applied Econometrics: A Modern Approach*, 4th ed. London, England: Red Globe Press, 2021.

[11] K.A. Nelson, R.J. Davis, D.R. Lutz, and W. Smith, "Optical Generation of Tunable Ultrasonic Waves," *Journal of Applied Physics*, Vol. 53, No. 2, pp. 1144–1149, 2002, doi: 10.1063/1.329864.

[12] V. Ratnasari, S.H. Audha, and A.T.R. Dani, "Statistical Modeling to Analyze Factors Affecting the Middle-Income Trap in Indonesia Using Panel Data Regression,' *MethodsX*, Vol. 11, pp. 1–9, 2023, doi: 10.1016/j.mex.2023.102379.

[13] A. Indrasetianingsih, K. Wasik, "Model Regresi Data Panel untuk Mengetahui Faktor yang Mempengaruhi Tingkat Kemiskinan di Pulau Madura," *Jurnal Gaussian*, Vol. 9, No. 3, pp. 355–363, 2020, doi: 10.14710/j.gauss.9.3.355-363.

[14] Z. Sun and J. Gou, "Panel Regression Method Based on Stata Analysis of Impact of Technology Innovation and Human Capital on Enterprise Value," in *Proceedings of the 2022 2nd International Conference on Business Administration and Data Science (BADS 2022)*, 2023, pp. 1307–1312.

[15] Z. Martha, B. Susetyo, and M.N. Aidi, "Panel Data Regression Model for Case of Dengue Hemorrhagic Fever (DHF) in Bogor," *Global Journal of Pure and Applied Mathematics*, Vol. 12, No. 1, pp. 741–746, 2016.

[16] D.J. Gubler, "Dengue and Dengue Hemorrhagic Fever," *Clinical Microbiology Review*, Vol. 11, No. 3, pp. 480–496, 1998.

[17] B.H. Baltagi, *Econometric Analysis of Panel Data*, 3rd ed. West Sussex, England: John Wiley & Sons Ltd, 2005.

[18] D.C. Montgomery, E.A. Peck, and G.G. Vining, *Introduction to Linear Regression Analysis*, 5th ed. Hoboken, NJ, USA: John Wiley & Sons Inc, 2012.

[19] W.H. Greene, *Econometric Analysis*, 5th ed. Upper Saddle River, New Jersey, USA: Prentice Hall., 2003.

[20] A.C. Rencher and G.B. Schaalje, *Linear Models in Statistics*, 2nd ed. Hoboken, NJ, USA: John Wiley & Sons Inc, 2008.

[21] T. Amemiya, "The estimation of the variances in a variance-components model," *International Economic Review*, Vol. 12, No. 1, pp.1–13, 1971, doi: 10.2307/2525492.

[22] C. Hsiao, *Analysis of Panel Data*, 2nd ed. Cambridge, United Kingdom: Cambridge University Press, 2003.