

Data Panel Modelling with Fixed Effect Model (FEM) Approach to Analyze the Influencing Factors of DHF in Pasuruan Regency

Achmad Faridz Jauhari¹, Yoyon K. Suprpto¹ and Achmad Mauludiyanto¹

¹Department of Electrical Engineering Institut Teknologi Sepuluh Nopember Surabaya, Indonesia

Keywords: Panel Data, Fixed Effect Model, Regression Analysis, Ordinary Least Square, Dengue Hemorrhagic Fever.

Abstract: Dengue Hemorrhagic Fever (DHF) is one of the endemic diseases caused by the bites of *Aedes* mosquitoes which are infected with the dengue virus. This disease can cause death. The DHF mortality rate in Pasuruan Regency is high (above 1% per year) in the last four years. Therefore, this study aimed to find a model that can explain the influencing factors of DHF incidence. The variables used were the number of DHF patients (Y), waste volume (X_1), rainy days (X_2), health facilities (X_3), temperature (X_4), number of high-educated population (X_5), population density (X_6), and rainfall (X_7). The data used were ranging from 2015 to 2018 and obtained from several agencies in Pasuruan Regency. In this study, the method used was the Panel Data Regression with Fixed Effect Model approach. The results of the model showed R^2 : 0.804 meaning that the seven variables were able to explain the effect on the incidence of DHF by 80.4% while the remaining 19.6% was influenced by other unknown variables. Of the seven predictor variables, there are six variables that have a significant effect consist of Waste Volume, Health Facilities, Temperature, Number of High-Educated Population, Population Density, and Rainfall. Henceforth, future DHF prevention and control policies can be more emphasized on these factors.

1 INTRODUCTION

Dengue hemorrhagic fever (DHF) is a disease caused by the dengue virus and is transmitted by *Aedes aegypti* and *Aedes albopictus* mosquitos. Both types of mosquitos are found in all corners of Indonesia, except in places with altitudes above 1000 masl (meters above sea level) (Arsin, 2013). The symptoms patients is a high fever for 2-7 days (38-400°C). At the acute level, this disease can cause death. Besides that, DHF can appear throughout the year and can affect all age groups (Kemenkes, 2017).

Indonesia is the country which has the highest DHF cases in Southeast Asia (Kemenkes, 2010). In the last five years, the highest number of dengue cases was occurred in 2016 reaching up to 204,171 cases with 1,598 deaths. The number of dengue cases in 2016 increased by 57.5% compared to the number of dengue cases in 2015 which was 129,650 cases. The number of DHF deaths in 2016 also worsens by 49.2% compared to the number of deaths in 2015 which was only 1,071 people. It was also reported that the Incident Rate (IR) in 2015 increased from 50.75 to 78.85 per 100,000 population. However, the Case Fatality Rate (CFR) has decreased from 0.83%

in 2015 to 0.78% in 2016. DHF has spread in 34 provinces and 463 districts/cities in Indonesia (Kemenkes, 2017).

Pasuruan Regency is one of the regions in East Java where its DHF mortality rate is above 1% in this past four years. Since 2015, DHF is determined as an Extraordinary Situation and as a result, the efforts to control DHF were carried out intensively. Based on the data released by the Public Health Office of Pasuruan Regency, the number of DHF cases in 2015 was 686 cases with 28 deaths whereas, in 2016, the number of DHF cases increased by 11% to 764 cases with a total death of 27 people. In 2017, the number of DHF cases decreased by 59% to 317 cases with 13 deaths and continued to decline by 38% with four deaths in 2018. The Incident Rate (IR) of DHF in 2015 was 43.31 per a 100,000 population while the Case Fatality Rate (CFR) was 4.1%. In 2016, the IR increased to 48.23 per a 100,000 population while the CFR dropped to 3.5%. On the other hand, in 2017, the IR decreased to 19.75 per 100,000 population while the CFR increased to 4.1%. In 2018, the IR was 11.90 per 100,000 population while the CFR was 2.1%. Based on the above data, the mortality/Case Fatality Rate of DHF in Pasuruan for the past four

years is included in the high category (CFR \geq 1%) (Kemenkes, 2017).

The spread of DHF in Indonesia is influenced by many factors such as climate, community behaviour, environment, as well as demographic and socioeconomic conditions (Arsin, 2013). The climate factors include temperature, rainfall, rainy days, humidity, and wind speed; behavioural factor is basically a person's response towards stimuli that is related to illness and disease including knowledge, actions, and beliefs associated with DHF, availability of health resources, adequate and affordability of health facilities, community support, as well as the concern from government and health workers; environmental factors are the number of places or containers of DHF vector proliferation; demographic factors consist of the population density and population mobility; socio-economic factors are the level of education, employment, and the number of family members. Of the factors mentioned above, knowing which of the factors that have a significant effect on DHF is very important so that future efforts to prevent and control DHF can be done more effectively and efficiently.

Modelling is a method that can be used to determine the factors that have a significant effect on DHF in Pasuruan Regency. This study aimed to find a model that is able to explain the influencing factors of DHF in 21 sub-districts of Pasuruan within the period of 2015 to 2018. Different regional characteristics in each sub-district are thought to influence the incidence of DHF. By that, this study used Panel Data structure. There has never been a research which examined the factors of DHF in Pasuruan Regency so that the results of this modelling are expected to be used as early warnings or basis for the formulation of strategic policies to prevent and eradicate DHF in the future.

Several studies related to the analysis of DHF have been done before. Research conducted by (Ariani, 2018) used the Negative Binomial regression model to produce variables that significantly influence DHF, namely population density, number of health workers and rainfall. Subsequent research by (Rustiani, 2017) using multiple linear regression resulted in R^2 values of 67 percent, significant variables included population density, rainfall and larva free index. Research conducted by (Rasmanto, 2016) using Linear Regression resulted in R^2 values of 24.1 percent, variables that had a significant effect consisted of air temperature and relative humidity. The study by (Hasirun, 2016) used a Spatial regression model with spatial error resulting in R^2 values of 43.34 percent, significant variables covering rainfall, health facilities, the percentage of PHBS houses, and the percentage

of healthy houses. Subsequent research by (Martha, 2015) used panel data regression with a fixed effect model approach resulted in R^2 values of 72,76 percent, significant variables included population density, population mobility, the average age of DHF patients, and the number of DHF patients in the previous time period. Based on these studies we use several variables that have been used previously namely population density, rainfall, rainy days, temperature, and the number of health facilities. Then we add two new variables namely the waste volume and the Number of High-Educated Population.

2 LITERATURE REVIEW

2.1 Panel Data

Panel Data is a collection of data consisting of cross-section data and time series data. Time series data usually includes one individual observed in a certain period of time while cross-section data is obtained from several individuals observed in a certain period of time. Therefore, it can be said that the Panel Data presents larger and more informative data and generates a greater degree of freedom. The general form of Panel Data regression models can be seen in Equation 1 (Hsiao, 2003; I. et al., 2017).

$$\gamma_{it} = \alpha_{it} + \beta' X_{it} + \varepsilon_{it} \quad (1)$$

- γ_{it} : Response variable of the i -individual and t -time period.
- α_{it} : Constant value/interception of i -individual and t -time period
- β' : $(\beta_1, \beta_2, \dots, \beta_k)$ Regression or slope coefficient variable of k -predictor
- X_{it} : The predictor variable value of the i -individual and t -time period
- ε_{it} : The error of i -individual and t -time period
- i : $1, 2, \dots, N$ number of individuals
- t : $1, 2, \dots, T$ time periods

2.2 Estimation of Panel Data Regression Model

In making parameter estimation, the model depends on the assumptions about intercepts and slope coefficients. Using Panel Data allows different intercepts and slope coefficients to occur in each individual and each time period. In this concept, there are three approaches that can be used, namely CEM (*Common Effect Models*), FEM (*Fixed Effect Model*) and REM (*Random Effect Models*) (Widarjono, 2013; I. et al., 2017)

2.2.1 CEM (Common Effect Models)

The approach with CEM model assumes that intercepts and slope coefficients have the same value for all individuals at all time periods. In other words, this model ignores individual dimensions and time. The equation of CEM model is illustrated in this following Equation 2.

$$\gamma_{it} = \alpha + \beta' X_{it} + \epsilon_{it} \tag{2}$$

2.2.2 FEM (Fixed Effect Model)

This approach believes that differences in characteristics between individuals are represented in the intercepts. Thus, intercepts for each individual will be different but the slope coefficients remain constant at all time periods. Equation 3 below presents the equation of FEM model.

$$\gamma_{it} = \alpha_i + \beta' X_{it} + \epsilon_{it} \tag{3}$$

2.2.3 REM (Random Effect Models)

In REM, the differences in individual characteristics are accommodated by error terms. Error terms may correlate between individuals and between times. For more details, the equation of REM model is formulated in Equation 4.

$$\gamma_{it} = \alpha + \beta' X_{it} + W_{it} \tag{4}$$

α is the mean of intercept from error terms cross section and time series. $W_{it} : \mu_i + \epsilon_{it}$ where μ_i is the random error cross section deviation, which explains the differences between individuals and each other.

2.3 Selection of Panel Data Regression Model

In choosing the model that will be used to manage Panel Data, several tests need to be done, namely:

2.3.1 Chow Test

Chow Test is used to choose a better model between CEM and FEM (Greene, 2000). The hypothesis in the Chow Test can be written as follows.

$$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_n = 0 \text{ (CEM)}$$

$$H_1 : \text{minimum of one } \alpha_i \neq 0 \text{ (FEM)}$$

Next, the equation for Chow Test Statistics is:

$$F_{count} = \frac{(RRSS - URSS)/(N - 1)}{URSS/(nT - n - K)} \tag{5}$$

Description: RRSS: *restricted residual sum of square* CEM URSS: *unrestricted residual sum of square* FEM n: number of individuals(*cross section*) T: time periods (*Time Series*) If the value of $F_{count} \geq F_{table} = F_{(n-1, nT-n-K)}$ or $p - value \leq 0.05$, then hypothesis H_0 is rejected. This means that FEM is a better model than CEM.

2.3.2 Hausman Test

Hausman test is used to choose a better model between FEM and REM (Baltagi, 2010; Sutikno et al., 2017). The hypothesis in the Hausman Test is as follows.

$$H_0 : \text{correlation}(\mu_i, X_{it}) = 0 \text{ (REM)}$$

$$H_1 : \text{correlation}(\mu_i, C_{it}) \neq 0 \text{ (FEM)}$$

This following equation 6 displays the Hausman Test Statistics.

$$W = (\hat{\beta}_{FEM} - \hat{\beta}_{REM})' [var(\hat{\beta}_{FEM} - \hat{\beta}_{REM})]^{-1} (\hat{\beta}_{FEM} - \hat{\beta}_{REM}) \tag{6}$$

If the value of $W \geq \chi_{(a,K)}^2$ or $p - value \leq 0.05$, then hypothesis H_0 rejected, meaning that FEM is a better model than REM. But if the value of $W < \chi_{(a,K)}^2$ or $p - value > 0.05$, then hypothesis H_0 , is accepted. This points out that the best model between FEM and REM is REM.

2.3.3 Lagrange Multiplier Test

LM Test is used to choose a better model between CEM and REM. LM Test does not need to be done if FEM is found to be the best in the Chow Test and Hausman Test. This test is only done when the best model found in the Chow Test is FEM whereas, in the Hausman Test, the best model is REM. These are the hypothesis in the LM test:

$$H_0 : \sigma_{\epsilon}^2 = 0 \text{ (CEM)}$$

$$H_1 : \sigma_{\epsilon}^2 \neq 0 \text{ (REM)}$$

In details, the statistics of LM Test can be understood through Equation 7 (Widarjono, 2013) below.

$$LM = \frac{nT}{2(T - 1)} \left(\frac{\sum_{i=1}^n (T \hat{\epsilon}_i)^2}{\sum_{i=1}^n \sum_{t=1}^T \hat{\epsilon}_{it}^2} - 1 \right)^2 \tag{7}$$

Description:

n: number of individuals

T: time period

$\hat{\epsilon}$: error on CEM

Hypothesis H_0 is rejected if the value of $LM \geq \chi_{table}^2 = \chi_{(a,K)}^2$ or $p - value \leq 0.05$. This indicates that REM is a better model than CEM. Otherwise, if the value of $LM < \chi_{table}^2 = \chi_{(a,K)}^2$ or $p - value > 0.05$, then the hypothesis H_0 is accepted which means that the correct model between the two models is CEM.

2.3.4 Classical Assumptions Test

To get the estimator of the parameter model in the CEM and FEM approach, Ordinary Least Square (OLS) method is used. The classical assumption is done to get the best OLS estimator known as BLUE (Best Linear Unbiased Estimator). Classical assumption tests include 1. Multicollinearity Test, 2. Normality Test, 3. Heteroscedasticity Test, and 4. Autocorrelation Test. From all the four tests, only test 3 and 4 that must be met to obtain the best estimator (BLUE) (Gujarati, 2004).

2.3.5 Multicollinearity Test

One of the assumptions used in OLS is the absence of a strong relationship between predictor variables. The presence of multicollinearity makes the model to have a large variant (Widarjono, 2013)). However, multicollinearity can still produce BLUE estimators. One way to detect multicollinearity is to use the sample of correlation coefficient (Gujarati, 2004; Sutikno et al., 2017). The sample of correlation coefficient is shown in the next Equation 8.

$$r_{12} = \tag{8}$$

$$\frac{NT \sum_{i=1}^N \sum_{t=1}^T x_{1it} x_{2it} - (\sum_{i=1}^N \sum_{t=1}^T x_{1it}) (\sum_{i=1}^N \sum_{t=1}^T x_{2it})}{\sqrt{[NT \sum_{i=1}^N \sum_{t=1}^T x_{1it}^2 - (\sum_{i=1}^N \sum_{t=1}^T x_{1it})^2] [NT \sum_{i=1}^N \sum_{t=1}^T x_{2it}^2 - (\sum_{i=1}^N \sum_{t=1}^T x_{2it})^2]}}$$

If the correlation coefficient is > 0.85 , then it is suspected that multicollinearity occurred in the model. Conversely, if the correlation coefficient is ≤ 0.85 , it is assumed that the model did not contain multicollinearity (Widarjono, 2013).

2.3.6 Normality Test

This assumption test is optional, but it is better if included in an effort to get a good OLS estimator. The hypotheses in this test are:

- $H_0 : E(\epsilon_{it}) = 0$ (error, distributed normally)
- $H_1 : E(\epsilon_{it}) \neq 0$ (error, not distributed normally)

The Normality Test Statistics adopted the Jarque-Bera formula as we can see in this following Equation 9 (Gujarati, 2004).

$$JB = NT \left(\frac{S_k^2}{6} + \frac{(K_r - 3)^2}{24} \right) \tag{9}$$

- Description,
- S_k : slope (Skewness)
- K_r : height (kurtosis)

If the value of $JB < \chi_{table}^2 = \chi_{(a,2)}^2$ or $p - value > 0.05$, then hypothesis H_0 is accepted. This clarifies that the error is normally distributed.

2.3.7 Heteroscedasticity Test

The assumption that the error variant has a constant (homoscedastic) nature is very important to produce the best OLS estimator. The Heteroscedasticity Test hypothesis is:

- H_0 error does not contain heteroscedasticity
- H_1 error contain heteroscedasticity

Glejser Method is used for the Heteroscedasticity Test Statistics, which is by regressing the absolute value of the error with its independent variable (Widarjono, 2013). If β_1 is not significant in t-test or the Prob value is > 0.05 , then hypothesis H_0 is accepted. This explains that the error did not contain heteroscedasticity or was constant.

2.3.8 Autocorrelation Test

This assumption test is used to see if there is no serial correlation on the error. This test is crucial to get a BLUE OLS estimator. One way to detect autocorrelation is to use a Durbin Watson test method. These are the hypothesis for the autocorrelation test in this study:

- $H_0 : Cov(\epsilon_{it}, \epsilon_{i,t-1}) = 0$ (no autocorrelation)
- $H_1 : Cov(\epsilon_{it}, \epsilon_{i,t-1}) \neq 0$ (there is autocorrelation)

The statistics of the Durbin Watson autocorrelation test can be seen in equation 10 (Gujarati and P, 2010).

$$DW = \frac{\sum_{i=1}^N \sum_{t=2}^T (\epsilon_{it} - \epsilon_{i,t-1})^2}{\sum_{i=1}^N \sum_{t=2}^T \epsilon_{it}^2} \tag{10}$$

Figure 1 below presents the critical value in the autocorrelation test:

$n < dw < d_L$: There is a positive autocorrelation
$d_L \leq dw \leq d_U$: There is no decision
$d_L \leq dw \leq 4-d_U$: There is no positive/negative: autocorrelation
$4-d_U \leq dw \leq 4-d_L$: There is no decision
$4-d_L \leq dw \leq 4$: There is a negative autocorrelation

Figure 1: Autocorrelation critical value of durbin watson justified.

2.3.9 Significance Test of Regression Parameters

The parameter significance test is used to find out the level of influence that the predictor variable has on the response variable. There are two tests carried out, namely simultaneous test and partial test.

2.3.10 Simultaneous Test

A simultaneous test is applied to determine the effect of predictor variables on the response variable together. The simultaneous test hypothesis is as follows.

$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$ (all predictor variables have no effect on the response variable)

$H_1 : \text{minimum of one } \beta_k \neq 0$ (there is at least one predictor variable that affects the response variable)

$$F_{count} = \frac{R^2 / (n + K - 1)}{(1 - R^2) / (nT - n - K)} \quad (11)$$

R^2 is the coefficient of determination. If the value of $F_{count} > F_{table} = F_{(a, n+K-1, nT-n-K)}$ or $p - value \leq 0.05$, then hypothesis H_0 , is rejected, meaning that in the model there is at least one predictor variable which has a significant effect on the response variable.

2.3.11 Partial Test

A partial test is used to determine the effect of each predictor variable on the response variable. The partial test hypothesis is:

$H_0 : \beta_k = 0$ (k -predictor variable does not significantly influence the response variable)

$H_1 : \beta_k \neq 0$ (k -predictor variable significantly influences the response variable)

The statistics of the partial test is formulated in Equation 12.

$$t_{count} = \frac{\beta_k}{se(\beta_k)} \quad (12)$$

If the value of $t_{count} > t_{table} = t_{(a/2, nT-n-K)}$ or $p - value \leq 0.05$, then hypothesis H_0 is rejected. This describes that k -predictor variable has a significant effect on the response variable.

3 LITERATURE REVIEW

In this research, the data used are secondary data obtained from several agencies in Pasuruan Regency including Department of Health, Department of Population and Civil Registration, Department of Environment, Department of Irrigation, and Meteorological, Climatological, and Geophysical Agency. The scope of the research is limited to 21 sub-districts in Pasuruan Regency which are at altitudes below 1000 masl in a span of four years (2015-2018). The steps of analysis can be seen in Figure 2.

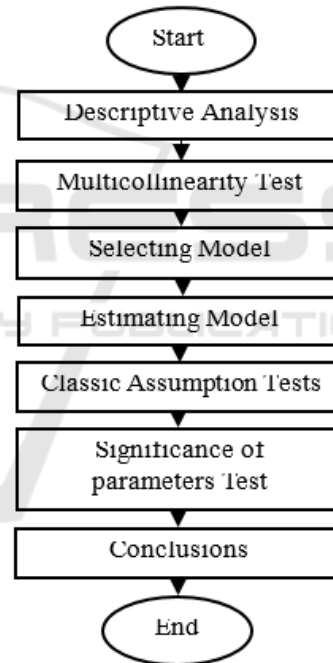


Figure 2: Steps of analysis.

Based on Figure 2, The steps of analysis in this study can be described as follows:

1. Conducting a descriptive analysis of the variables used in the study.
2. Performing multicollinearity test on all predictor variables (X) using the sample correlation coefficient.
3. Selecting the Panel Data regression model by conducting the Chow Test, Hausman Test, and Lagrange Multiplier Test

4. Estimating the Panel Data regression model with the approach chosen in step 3.
5. Carrying classic assumption tests which include normality test, Heteroscedasticity test, and autocorrelation test.
6. Testing the significance of the regression parameters through Simultaneous Test (f-test) and Partial Test (t-test).
7. Making conclusions and suggestions.

The variables which are suspected of having an effect on DHF are presented in Figure 3.

Var.	Description	Unit
Y	Number of DHF patients	people
X ₁	Volume of untransported waste	dam ³
X ₂	Number of rainy days	day
X ₃	Number of health facility	unit
X ₄	Temperature	°C
X ₅	High educated people	people per a 10,000 population
X ₆	Population density	people/ km ²
X ₇	Rainfall	mm

Figure 3: Variables assumed to influence dhf.

4 RESULTS AND DISCUSSION

4.1 Descriptive Analysis

The first step in this research is to do a descriptive analysis that is useful to know the characteristics of each research variable. Descriptive analysis used includes average, maximum, and minimum values. The results of the descriptive analysis are shown in Figure 4.

Variable	Mean	Max.	Min.
Y	23	115	0
X ₁	54.26	95.12	11.04
X ₂	110	161	71
X ₃	53	135	13
X ₄	26.32	27.23	22.96
X ₅	253	712	75
X ₆	1,631	2,954	325
X ₇	1,640.41	3,309	707

Figure 4: Descriptive statistics of research variables.

Based on Figure 4, the average number of DHF patients (Y) in 21 sub-districts of Pasuruan in this past four years (2015-2018) is 23 people. Lumbang and Puspo sub-districts have the least number of patients (none) while Bangil has the highest number of patients (115 people). The average waste (X₁) that is not transported is 54.26 where Kraton becomes the sub-district that has the most non-transported waste reaching up to 95.12 dam³. In contrast, Bangil has the least non-transported waste which is only 11.04 dam³. The range of maximum and minimum values of the waste volume that is not transported is very

high. This indicates that the waste services are still focused on certain sub-districts or in other words, not comprehensive in all sub-districts. The average number of rainy days (X₂) is 110 where Purwosari has the most number of rainy days reaching up to 161 days of rain in a year. Meanwhile, the sub-district with the least number of rainy days is Gondangwetan and Winongan, 71 days a year. On the other hand, the average number of health facilities (X₃) is 53. The highest number of health facilities is owned by Pandaan sub-district (135 units) while the least health facilities are owned by Puspo (13 units). The average temperature (X₄) is 26.32°C. It is reported that the highest temperature occurs in Beji reaching up to 27.23°C while the lowest temperature is in Purwodadi reaching up to 22.96°C. The average number of educated population (X₅) is 253 where the area with the highest number of educated people is Bangil (712 people per a 10,000 population) while the lowest number of educated people is Lekok (75 people per a 10,000 population). The average population density (X₆) is 1,631. The most densely populated area is Pohjentrek with a population density of 2,954 people/km² whereas the lowest population density can be found in Lumbang, only 325 people/km². The range between the maximum and minimum population density is very big. This means that the population distribution is uneven or concentrated in certain sub-districts. Last but not least, the average value of Rainfall (X₇) is 1,640.41 where the area with the highest rainfall intensity is Purwosari reaching up to 3,309 mm in a year. Meanwhile, the area with the lowest rainfall intensity is Winongan, only 707 mm in a year.

4.2 Multicollinearity Test

Multicollinearity test is applied to see the correlation of each predictor variable used in the regression model. The existence of perfect multicollinearity can cause many predictor variables do not have a significant effect although the coefficient of determination is high. To detect multicollinearity, the correlation coefficient method is used (r). The results of multicollinearity test are presented in Figure 5.

	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇
X ₁		0.01	-0.09	0.09	-0.42	-0.05	0.01
X ₂	0.01		0.23	-0.34	0.16	-0.08	0.84
X ₃	-0.09	0.23		0.31	0.76	0.54	0.27
X ₄	0.09	-0.34	0.31		0.20	0.55	-0.46
X ₅	-0.42	0.16	0.76	0.20		0.60	0.27
X ₆	-0.05	-0.08	0.54	0.55	0.60		-0.06
X ₇	0.01	0.84	0.27	-0.46	0.27	-0.06	

Figure 5: Correlation coefficient.

From Table 4, the value of r in each column is < 0.85 so that all predictor variables are free from multicollinearity problems.

4.3 Selection of Panel Data Regression Model

To choose a model that fits the research data, researchers can conduct the Chow Test, Hausman Test, and Lagrange Multiplier Test. Chow Test is used to determine whether or not the differences between the characteristics of the districts are seen from the intercept on each regression model. From the Chow Test, it is obtained that $F_{count} = 5.618 > F_{0.05(20;56)} = 1.76$ dan $p - value = 0.0000 \leq 0.05$ with a level of reliability by 95%. This means that hypothesis H0 rejected, there are differences in the regional characteristics presented in the DHF Panel Data regression equation model of Pasuruan Regency so that the appropriate model must be FEM.

Moreover, the Hausman Test is performed to test whether or not there is a correlation between the error component in the model and its predictor variable. The Hausman Test results showed that $W_{count} = 79.223 > \chi^2_{(005;7)} = 14.07$ and $p - value = 0.0000 \leq 0.05$ with a level of reliability by 95%. It can be said that the hypothesis H0 is rejected. This points out that there is no correlation between the error component predictor variables on Pasuruan DHF Panel Data regression equation. Therefore, the corresponding model is FEM.

Because of the two tests above refer FEM as the appropriate model, it is not necessary to do the Lagrange Multiplier Test.

4.4 Model Estimation with FEM Approach

Here is the estimation of Panel Data regression model formed by using FEM approach.

$$\hat{y}_{it} = \hat{\alpha}_i + 1.785X_{1it} + 0.186X_{2it} + 1.470X_{3it} + 32.267X_{4it} + 3.186X_{5it} - 0.177X_{6it} - 0.014X_{7it} \quad (13)$$

$\hat{\alpha}_i$ is the intercept that has different values in each region. This value will distinguish the prediction of the number of DHF patients in between sub-districts. Because each region has different characteristics, the characteristic differences are represented by an intercept variable in the *fixed effect* Panel Data regression model. The intercept value for each district is shown in Figure 6.

Equation 13 shows that the increase in variable Y is influenced by the increase in variable $X_1, X_2, X_3, X_4,$ and X_5 . Meanwhile, the increase in variable X_6 and X_7 will affect the decrease in variable Y.

Index (i)	District	Intercept ($\hat{\alpha}_i$)
1	Kejayan	-1,335.551
2	Wonorejo	-1,276.405
3	Gempol	-1,868.230
4	Beji	-1,910.659
5	Bangil	-2,840.675
6	Rembang	-1,142.526
7	Kraton	-1,273.838
8	Pohjentrek	-1,364.992
9	Gondangwetan	-1,683.950
10	Rejoso	-1,537.214
11	Winongan	-1,322.926
12	Grati	-1,599.886
13	Lekok	-925.545
14	Nguling	-1,184.983
15	Lumbang	-1,096.910
16	Pasrepan	-1,132.060
17	Purwosari	-1,734.530
18	Sukorejo	-1,662.592
19	Pandaan	-2,404.691
20	Purwodadi	-1,537.738
21	Puspo	-1,027.002

Figure 6: Estimated intercept value of each region.

4.5 Classical Assumptions Test

After the estimation of the Panel Data regression model is obtained, the next step is to carry out the classical assumption tests which include the Normality Test, Heteroscedasticity Test, and Autocorrelation Test.

The Normality Test is done by using the Jarque-Bera formula. From this test, it is obtained $JB = 0.372 < \chi^2_{(005;2)} = 5.991$ and $p - value = 0.83 > 0.05$ with a level of reliability by 95%. By that, hypothesis H_0 is accepted meaning that the error is normally distributed. Then, a Heteroscedasticity Test is performed to see whether the error variant has a constant nature or not. Glejser Test is used for the Heteroscedasticity Test and the values obtained can be seen in Figure 7 below.

Variable	t_{count}	$p - value$
X_1	-1.355	0.181
X_2	0.387	0.700
X_3	0.671	0.505
X_4	-0.811	0.421
X_5	-0.015	0.989
X_6	-1.990	0.052
X_7	0.345	0.731

Figure 7: Simultaneous test result.

Based on Figure 7, the p-value for each predictor variable is > 0.05 making the hypothesis H_0 accepted. This explains that the *error* in the model does not contain heteroscedasticity or is constant.

The last classical assumption test is the Autocorrelation Test that is done by using the Durbin Watson test method. It is obtained that $DW = 2.056$.

The value is in the area of Hypothesis H_0 which is $d_u = 1.8291 \leq DW = 2.101 \leq 4 - d_u = 2.1709$, meaning that the *error* in the model is free from the problem of autocorrelation.

4.6 Significance of Parameters Test

After the model has passed all classical assumption tests, the next step is to do a simultaneous test and partial test to determine the effect of predictor variables on the response variable both simultaneously and partially. The simultaneous test results are presented in Figure 8.

R-squared	0.8044
S.E. of regression	11.9153
Sum squared resid	7950.5320
F-statistic	8.5279
Prob(F-statistic)	0.0000

Figure 8: Simultaneous test result.

As shown in Figure 8, the value of $F_{count} = 8.528 > F_{(0.05;27;56)} = 1.686$ and $p - value = 0.000 < 0.05$ and $R^2 = 0.804$ with a level of reliability by 95%. This clarifies that hypothesis H_0 is rejected, the predictor variable simultaneously affects the response variable. The value of $R^2 = 0.804$ shows that the seven predictor variables can affect the number of DHF patients in Pasuruan by 80.4% while the remaining 19.6% are influenced by other variables outside the model.

After that, a partial test is carried out. The results of the partial test are shown in Figure 9.

Variable	t_{count}	$p - value$
X1	2.271	0.027
X2	1.487	0.143
X3	2.730	0.009
X4	3.076	0.003
X5	2.732	0.008
X6	-2.463	0.017
X7	-2.237	0.029

Figure 9: Partial test results.

As presented in Figure 9 with a reliability level of 95%, there are six variables that have a significant effect on the response variable ($p - value < 0.05$), namely variable $X_1, X_3, X_4, X_5, X_6,$ and X_7 while the variable X_2 does not have a significant influence on the response variable ($p - value > 0.05$).

5 CONCLUSION AND SUGGESTION

Based on the results of the analysis, it can be concluded that the Panel Data regression model with Fixed Effect Model approach can explain the effect of

predictor variables on the response variable. Therefore, the model equation is:

$$\hat{y}_{it} = \hat{\alpha} + 1.785X_{1it} + 0.186X_{2it} + 1.470X_{3it} + 32.267X_{4it} + 3.186X_{5it} - 0.177X_{6it} - 0.014X_{7it} \quad (14)$$

From the Equation 14 above, the value of $R^2 = 0.804$ shows that the percentage of all seven predictor variables able to affect the number of DHF patients in Pasuruan by 80.4% while the other 19.6% are influenced by other variables outside this study. Of the seven predictor variables, there are six variables that significantly influence the increase in the number of DHF patients such as waste volume, number of health facilities, temperature, high-educated population, population density, and rainfall.

Last but not least, the Department of Health of Pasuruan Regency is suggested to plan the programs and activities to control and eradicate future DHF by focusing on the variables that have a significant effect on DHF patients.

In the future the model can be developed into an application that can be utilized by government or interested organizations.

REFERENCES

Ariani, P. M. (2018). *Analisis Faktor yang Berpengaruh Terhadap Pencegahan Penyakit DBD di Provinsi Jawa Tengah Menggunakan Regresi Binomial Negatif. Tugas Akhir.* Universitas Islam Indonesia, Yogyakarta.

Arsin, A. A. (2013). *Epidemiologi demam berdarah dengue (DBD).* Masagena Press, di indonesia.Makassar.

Baltagi, B. H. (2010). *Econometric Analysis of Panel Data.* John Wiley & Sons, Ltd, Sussex, 4th edition. west edition.

Greene, W. (2000). *Econometric Analysis.* Prentice Hall, Inc, 4th edition.

Gujarati, D. (2004). *Basic Econometrics.* The McGraw-Hill Companies, New York, 4th edition.

Gujarati, D. and P, D. C. (2010). *Dasar-dasar Ekonometrika Edisi 5.* Salemba Empat, Jakarta.

Hasirun (2016). *Model Spasial Faktor Risiko Kejadian Demam Berdarah Dengue di Provinsi Jawa Timur Tahun 2014. Tesis.* Universitas Airlangga, Surabaya.

Hsiao, C. (2003). *Analysis of Panel Data.* Cambridge University Press, New York.

I., A. W., Ratnasari, V., and Wibowo, W. (2017). Analisis faktor yang berpengaruh terhadap tingkat pengangguran terbuka di provinsi jawa timur menggunakan regresi data panel. *Jurnal Sains dan Seni ITS*, 6(1).

Kemenkes, R. (2010). *Jendela Epidemiologi.* Pusat Data dan Surveilans Epidemiologi Kementerian Kesehatan Republik Indonesia, Jakarta.

- Kemenkes, R. (2017). *Profil Kesehatan Indonesia Tahun 2017*. Setjen Kementerian Kesehatan Republik Indonesia, Jakarta.
- Martha, Z. (2015). *Pemodelan Regresi Data Panel Pada Kasus Jumlah Penderita Demam Berdarah Dengue (DBD) di Kota Bogor.Tesis*. Institut Pertanian Bogor, Bogor.
- Rasmanto, M. (2016). Model prediksi kejadian demam berdarah dengue (dbd) berdasarkan unsur iklim di kota kendari tahun 2000-2005. *Jurnal Ilmiah Mahasiswa Kesehatan Masyarakat*, 1.
- Rustiani, N. N. (2017). Pemodelan penyebaran penyakit demam berdarah dengue (dbd) dengan pendekatan regresi linear berganda. In *Prosiding Seminar Nasional Sainstek 2017*. Fakultas MIPA Universitas Udayana.
- Sutikno, B., F.k, A., and D, O. (2017). Penerapan regresi panel data komponen satu arah untuk menentukan faktor-faktor yang mempengaruhi indeks pembangunan manusia.jurnal matematika integratif unpad.
- Widarjono, A. (2013). *Ekonometrika Teori dan Aplikasi Untuk Ekonomi dan Bisnis*. Ekonosia FE UII, Yogyakarta.

