# Robust Regression Estimation Mm Bisquare On The Factor of DHF In East Java

*by* 05 Mardiana Mardiana

# Robust Regression Estimation Mm Bisquare On The Factor of DHF In East Java

**Mardiana, M.Kes**[1*], Sri Sunarti[2], Vita Pramaningsih[3], Chatarina U.W [4], Lutfi Agus S[5]

[1]*Bachelor Degree of Public Health Department, Faculty of Health and Pharmacy, Muhammadiyah University of East Borneo Indonesia*
[2]*Bachelor Degree of Public Health Department, Faculty of Health and Pharmacy, Muhammadiyah University of East Borneo Indonesia*
[3]*Diploma of Environmental Health Department, Faculty of Health and Pharmacy, Muhammadiyah University of East Borneo Indonesia*
[4]*Epidemiology Department, Master Degree of Public Health Department, Faculty of Public Health, Airlangga University Indonesia*
[5]*Biostatistics and Demography Department, Master Degree of Public Health Department, Faculty of Public Health*
**Corresponding author: mar348@umkt.ac.id*

## Abstract

Estimation MM is an alternative of OLS Regression on the data outlier. East Java is one of area that has CFR > 1% (high category). There are some factors that can affect IR DHF such as climate change, demography, community's behavior and environmental sanitation. This research aims to know the regression factors that affect IR DHF in the East Java in 2017 by using estimation MM Bisquare on the data outlier. This research has non-reactive (Unobtrusive) design by using secondary data. The independent variable is population density, percentage of PHBS, percentage of healthy home and also precipitation of East Java in 2017, while the dependent variable was IR DHF in 2017. The population is 38 regencies in East Java. Meanwhile, the sample is 35 regencies chosen by simple random sampling. The analysis of the data is using regression on estimation MM with Tukey's Bisquare. Regression estimation MM Bisquare is the effective regression method on the data outlier. The regression model is (y ) ?= 22.325 + 0.010 (population density) + 0.207 (% of PHBS) – 0.527 (% of healthy home) + 0.006 (precipitation) with R2 adjusted (0.522) and MSE (86.026). The density of population and percentage of healthy home (p value < 0,05) affect the IR DHF of East Java in 2017 on the significant standard 0.05, while the percentage of PHBS and precipitation (p value > 0.05) has no effect on the IR DHF in East Java in 2017. Regression estimation MM Bisquare used as the alternative regression on the data outlier. The factors that affect IR DHF can be the main focus of DHF prevention program for the government and community.

Keywords**:** DHF, Robust Regression, Estimation MM, Bisquare

04/01/2020

**Mardiana**
Bachelor Degree of Public Health Department Faculty of Health and Pharmacy
UMKT, Samarinda City, East Kalimantan Province, Indonesia

Co-authors: Sri Sunarti, Vita Pramaningsih, Chatarina Umbul Wahyuni, Lutfi Agus Salim

Dear Mardiana,

**ACCEPTANCE OF ARTICLE FOR PUBLICATION IN ANNNALS OF TROPICAL MEDICINE & PUBLIC HEALTH**

I am happy to inform you that your article titled: **"Robust regression MM-Bisquare estimation on the factor of DHF in East Java"** has been accepted for publication in Annals of Tropical Medicine and Public Health. Provisionally, it is scheduled to be published in the forthcoming January 2020 Special Issue.

Accept my congratulation for this. I look forward to receiving more articles from you.

Best regards,

**Assistant Professor Abubakar Yaro PhD**
**Editor In Chief**
Annals of Tropical Medicine & Public Health
Editorial Office
AHRO Scientific Publishing Ltd
Glasgow Scotland

# Robust regression MM-Bisquare estimation on the factor of DHF in East Java

Mardiana Mardiana[1*], Sri Sunarti[1], Vita Pramaningsih[2], Chatarina Umbul Wahyuni[3], Lutfi Agus Salim[4]

[1]Bachelor Degree of Public Health Department, Faculty of Health and Pharmacy, Muhammadiyah University of East Kalimantan, Indonesia
[2]Diploma of Environmental Health Department, Faculty of Health and Pharmacy, Muhammadiyah University of East Kalimantan, Indonesia
[3]Epidemiology Department, Master Degree of Public Health Department, Faculty of Public Health, Airlangga University Indonesia
[4]Biostatistics and Demography Department, Master Degree of Public Health Department, Faculty of Public Health

***Corresponding author:**
Mardiana Mardiana
Bachelor Degree of Public Health Department Faculty of Health and Pharmacy
UMKT, Samarinda City, East Kalimantan Province, Indonesia
Phone numbers: +6281351414752
E-mail address: mar348@umkt.ac.id

## ABSTRACT

**Background:** MM Estimation is an alternative of regression on the data outlier. East Java is one of area that has CFR > 1%. There are some factors that can affect IR DHF such as climate change, community behavior, and environmental. **Aims:** to know the regression factors that affects the IR DHF in East Java in 2017. **Settings and Design:** This research has non-reactive (unobtrusive) design by using secondary data. **Methods and Material**: The independent variables are population density, percentage of PHBS, percentage of healthy home, and precipitation of East Java in 2017. The dependent variable is IR DHF in 2017.The population is 38 regencies and cities in East Java with the sample are 35 regencies and cities by simple random. **Statistical analysis used:** Regression on MM estimation with Tukey's Bisquare. **Results:** Regression MM-Bisquare estimation is the effective on the data outlier. The regression model is $\hat{y}$ = 22.325 + 0.010 (population density) + 0.207 (% of PHBS) – 0.527 (% of healthy home) + 0.006 (precipitation). The density of population and percentage of healthy home (p value < 0,05) affect the IR DHF of East Java in 2017 on the significant standard 0.05, while the percentage of PHBS and precipitation (p value > 0.05) have no effect on the IR DHF in East Java in 2017.**Conclusions:** Regression MM-Bisquare estimation is used as the alternative regression on the data outlier. The factors that affect IR DHF can be the main focus of DHF prevention program.

**Keywords**: DHF, Robust Regression, MM Estimation, Bisquare

**How to cite this article: Mardiana M, Sunarti S, et al (2021): Robust regression MM-Bisquare estimation on the factor of DHF in East Java, Ann Trop Med & Public Health; 22(S01): SP24132. DOI: http://doi.org/10.36295/ASRO.2021.24132**

## Introduction

The analysis of linear regression is the analysis conducted to know about the correlation between some independent and dependent variables, as well as as a prediction (1). The smallest quadrate or Ordinary Least Square (OLS) method is a general method used for the estimation of parameter regression by minimizing the total residue (error) to get the parameter estimation. One of conditions that must be completed by the use of OLS is the normal residue distribution (2). On certain condition, the residue cannot be distributed normally because there is an outlier on the data (3). The OLS is sensitive about the data outlier. The parameter outcome is refraction (4).

Outlier is an extreme observation among others. If the outlier does not occur because of an error in the measuring process or the data entry, then the data that includes outlier cannot be removed to improve the agreement, but it will increase the risk of carefulness problem in the estimation (5). The outlier can give the information that has not been given by the other data; therefore, the data outlier is important. In this condition, it needs a robust method toward the data outlier by using robust regression (3).

The use of robust regression will produce robust or resistance model in the data outlier (6). Resistance estimation means that the estimation is relative and not influenced when having a big change on the small data or otherwise (7). Method of Moment (MM) Estimation is one of the estimators used in the robust regression(8). The criteria of robust estimator can be measured from the breakdown point and the efficiency of estimator (4). These two criteria can be found on the MM estimation, including high break downpoint up to 50% with 95% efficiency (8).

Nowadays, many sectors use the robust regression analysis model. One of them is in the health field modeling data, such as DHF cases. Some of the previous researches mentioned that the data cases of DHF tend to be an outlier, so it is needed a special analysis method to resolve the outlier (9). This disease is caused by dengue virus infected to human by *Aedes sp* (10). DHF is a public health problem that still happens in some countries in this world until now. World Health Organization (WHO) estimates that about 50-100 million people are infected dengue every year, including 500 thousand cases of DHF with 22 thousand fatalities. Asia is the first with 70% of population which has the risks, among others are the Southeast Asia and the West Pacific.

DHF has increased and spread wider than before in Indonesia (11). There were 59,047 cases (IR : 22,55% of 100,000 inhabitants) and a fatality as many as 444 (CFR : 0,75%) in 2017 (12). CFR (Case Fatality Rate) is an indicator of high and low DHF cases. If the percentage of CFR is > 1% , it means high category case (13). One of 34 provinces in Indonesia that still has DHF cases is East Java. Based on the Health Service of East Java in 2017, there have been as many as 7,866 cases of DHF (IR : 20% of 100,000 inhabitants) with 106 fatalities (CFR : 1,34%) of the 38 regencies in East Java (14).

Many factors can affect the height of dengue cases in an area, such as climate change that makes an increase in breeding *Aedes sp* andthe change of behavior and the condition of population (14). The government has still continued doing a variety of preventive efforts and the control of DHF to reduce the risk for inhabitants until now. Therefore, it needs to conduct research about modeling factors of the occurrence of DHF in East Java by using the data outlier of DHF cases.

## Subjects and Methods

This research is an observational research with non-reactive (unobtrusive) design. Non-Reactive research is the research focused on the data information of individual at the past which was recorded by the secondary data without awareness and disturbance. The collection of secondary data analysis of this research was conducted without involving the individuals directly. This research was conducted in 2019 and has passed the fit of ethics in The Ethical Commission of Studies Health FKM UNAIR with certificate numbered 111/EA/KEPK/2019.

The independent variables consist of population density, percentage of Clean and Healthy Living Behavior (PHBS), percentage of healthy home, and precipitation. The dependent variable is IR DHF of East Java in 2017. The population is the whole area in East Java as many as 38 regencies and cities. The sample is 35 regencies and cities chosen by simple random sampling. The identification of outlier used graphic method and Z score. The data analysis used linier regression on MM estimation using Tukey's Bisquare.

## Results

East Java is located around the equator in Indonesia. There are two seasons every year, namely dry and rainy seasons. East Java is 47,799.75 km$^2$ that consists of 29 regencies and 9 cities. The total population of East Java in 2017 was 39,292,972 people. That total population has increased up to 0.53% compared to that of in 2016, which was up to 39,075,152 people. There are three areas with the biggest population in 2017, namely Surabaya (2,874,699 people), Malang regency (2,576,596 people), and Jember regency (2,430,185 people) (15).

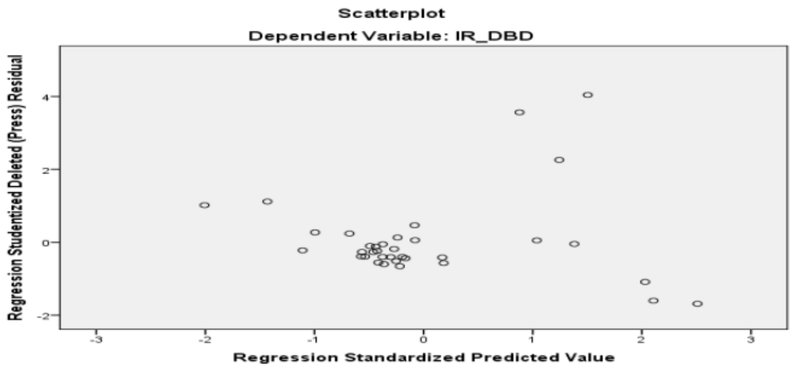
### a. The analysis of Regression OLS (MKT) and Outlier Identification

The first step before doing analysis of robust regression is conducting the analysis of linier regression using OLS. On the analysis of OLS linier regression, the case number of DHF (IR DHF) is the dependent variable, while the independent variables are population density, % of Clean and Healthy Living (PHBS),% of healthy home, and precipitation.

Based on the result of regression analysis by using OLS, the F value is 1,074 with *p value* up to 0.387. If compared with α, it was 0.05 and *p value* 0.387 > 0.05, so the $H_0$ is accepted. It means there is no influence of population density, % of PHBS, % of healthy home, and precipitation on the case number of DHF. Therefore, MSE produced was 2,056.321 with $R^2$ *adjusted* up to 0.009.

The analysis of the OLS regression will produce residue used for identification of outlier on the data by using graphic method and Z score. The first method of outlier identification is using graphic method approach. The residue on the research data will be visualized in the form of scatter plot on the graphic method.

**Figure 1**. **Residual Scatter Plot of Research Data**




Based on the residue plot on the **Figure 1,** generally, some data are far from the point of data. The point of data that spread was called outlier; therefore, based on the identification result by using the graphic method, it can be concluded there is an outlier on the observation.

Furthermore, the outlier identification is made in univariate by using Z score method, which is outlier identification in every observation on the independent and dependent variables. The data are identified outlier if they have the absolute Z score with standardization of > 3 theoretically.

**Table 1 : Results of Standarized Z Score**

| No. | Variables | District/City | \|Standarized Z Score\| |
|---|---|---|---|
| 1 | Incidence Rate | Blitar City | 3.6232 |
| 2 | Incidence Rate | Batu City | 3.02526 |
| 3 | Population Density | Surabaya City | 3.27558 |
| 4 | Percentage of Clean and Healthy Living Behavior (PHBS) | Ngawi Regency | 3.05624 |

Based on the **Table 1,** the outlier can be identified in univariate; there are four research data with the absolute Z score > 3, so it was indicated as data outlier. It can be concluded that the research data have an outlier on the dependent variable (vertical outlier) and independent variables (leverage point).

**b. The analysis of Robust Regression on MM-Bisquare Estimation**

The first weighting was by using the result of residual unstandardized from the regression agreement model with S estimation that has been convergent on the robust regression MM estimation. After passing the weighting process of S estimation (tunning constant = 1.547) until getting the convergent, the residual unstandardized will be used for weighting using M estimation (tunning constant = 4.685). The weighting using M estimation will be done until getting the convergent regression. The last convergent is the regression with MM estimation. The detailed steps of it are as follows:

1) Interprete $\beta$ that is $\hat{\beta}_0$ using S estimation, so it will get the breakdown pointup to 50%.

2) Count the residual $e_i = y_i - \hat{y}_i$

3) Count the total of $\hat{\sigma}_i = \frac{median|e_i - median(e_i)|}{0,6745}$.

4) Count the total of $u_i = \frac{e_i}{\hat{\sigma}_i}$

5) Count the value of weighting $W_i$ using Tukey'sBisquare with tunning constant $c$ = 4.685 so that it will get the efficiency up to 95%.

$$W(u,c) = \begin{cases} \left(1 - \frac{u^2}{4.685^2}\right)^2, & \text{if } |u| \leq 4.685 \\ 0, & \text{if } |u| > 4.685 \end{cases}$$

6) Count the total of $\hat{\beta}_{MM}$ by using OLS based on the total of $W_i$.

$\hat{\beta}_{MM} = (X'WX)^{-1}X'Wy$

7) Do the step (b) until (f) so that will get the convergent value of $\hat{\beta}_{MM}$ that is the total difference of $\beta_{MMj}^{l+1}$ and $\beta_{MMj}^{l}$ approximate to 0.

8) Do the significant test simultaneously (F test) with the hypothesis testings as follows :

$H_0$ = there is no significant effect between dependent and independent variables.

$H_1$ = there is significant effect between dependent and independent variables

With the rejection criteria of $H_0$ if p value is < 0.05

9) Do the significance partially (uji T) with the rejection criteria $H_0$ if p value is < 0.05. with hypothesis testings as follows :

   $H_0$ = there is no significant influence between independent and dependent variables.

   $H_1$ = there is significant influence between independent and dependent variables.

10) Get the modeling based on robust regression MM estimation.

Based on the result of robust regression analysis on MM estimation, it was obtained convergent estimation parameter after done in 16 times iteration. The iteration results of MM estimation are as follows:

**Table 2 : Tukey's Bisquare Weighted Estimation of MM, Iteration Results of Robust Regression Analysis**

| Iterasi | $B_0$ | $B_1$ | $B_2$ | $B_3$ | $B_4$ |
|---|---|---|---|---|---|
| 1. | 26,934 | + 0,011 | – 0,126 | - 0,507 | + 0,005 |
| 2. | 24,404 | + 0,011 | – 0,181 | - 0,531 | + 0,005 |
| 3. | 23,519 | + 0,011 | – 0,194 | - 0,531 | + 0,006 |
| 4. | 22,855 | + 0,010 | – 0,196 | - 0,524 | + 0,006 |
| 5. | 22,631 | + 0,010 | – 0,202 | - 0,527 | + 0,006 |
| 6. | 22,498 | + 0,010 | – 0,204 | - 0,527 | + 0,006 |
| 7. | 22,426 | + 0,010 | – 0,205 | - 0,527 | + 0,006 |
| 8. | 22,384 | + 0,010 | – 0,206 | - 0,527 | + 0,006 |
| 9. | 22,359 | + 0,010 | – 0,206 | - 0,527 | + 0,006 |
| 10. | 22,345 | + 0,010 | – 0,206 | - 0,527 | + 0,006 |
| 11. | 22,336 | + 0,010 | – 0,206 | - 0,527 | + 0,006 |
| 12. | 22,331 | + 0,010 | – 0,207 | - 0,527 | + 0,006 |
| 13. | 22,328 | + 0,010 | – 0,206 | - 0,527 | + 0,006 |
| 14. | 22,326 | + 0,010 | – 0,206 | - 0,527 | + 0,006 |
| 15. | 22,325 | + 0,010 | – 0,206 | - 0,527 | + 0,006 |
| 16. | 22,325 | + 0,010 | – 0,206 | - 0,527 | + 0,006 |

Source : Primary Data

Based on the **Table 2**, the analysis result of robust regression by using MM estimation (*Tukey's Bisquare*) was a convergent robust regression on the fifteenth iteration with the agreement of robust regression as follows:

$$\hat{y} = 22.325 + 0.010\ X_1 - 0.206\ X_2 - 0.527\ X_3 + 0.006\ X_4$$

The agreement of robust regression has the F value of 10,301 with *p value*of 0.000. If it compared with α up to 0.05, the p value was $0.000 < 0.05$. Therefore, $H_0$ was rejected. It means there is an influence of population density, % of PHBS, % of healthy home, and the precipitation on the number of DHF cases that has been rejected. Therefore, it can be concluded that based on the analysis of robust regression on MM estimation, it is obtained the population density, % of PHBS, % of healthy home, and precipitation which have significant influence on the number of DHF cases.

**Table 3 : Coefficient of Robust Regression Analysis Results of MM Estimation with Tukey's Bisquare Weight**

| Model | B | T Count | Significance | MSE | $R^2$ *adjusted* |
|---|---|---|---|---|---|
| Constant | 22.325 | 1.729 | 0.094 | | |
| Population Density | 0.010 | 5.589 | 0.000 | | |
| % Clean and Healthy Living Behavior (PHBS) | 0.207 | 1.712 | 0.097 | 86.026 | 0.522 |
| % Healthy Home | -0.527 | -4.463 | 0.000 | | |
| Precipitation | 0.006 | 1.386 | 0.176 | | |

Based on the **Table 3,** when the significant level α is up to 0.05 on the robust regression by using MM estimation with Bisquare, p value on the population density and % of healthy home is up to 0.000 < 0.05, so that $H_0$ is rejected. It means that the population density and the percentage of healthy home have significantly influenced the number of DHF cases. Every rising up to 1 people/km$^2$ will increase the IR DHF up to 0.010/100,000 people. It means that if there is a rising in population density up to 50 people/km$^2$, it will make a rising in IR DHF up to 0.5/100,000 people.

Moreover, every 1% derivation of healthy home will make a rising in IR DHF up to 0.527/100,000 people. It means if there is derivation of % of healthy home up to 50%, it will make arising in IR DHF is up to 26.35/100,000 people. Meanwhile, the percentage of PHBS and precipitation has p value > 0.05, so $H_0$ is accepted. It means the percentage of PHBS and precipitation has no significant influence on the number of DHF cases.

## Discussion

Generally, the linier regression used OLS in estimation of parameter regression, but in the data outlier, that method cannot work better because influenced by outlier (5). There are some deviatons that can happen on data outlier such as the residue from the mode and the variety of data that will become bigger (16). The analysis of data outlier will impact to the estimation parameter with refraction and bias interpretation. The removal of data outlier is not a right thing to do because the outlier includes the important thing that cannot be found on the other data in the certain exact condition, for example, the outlier that exists because of the combination of certain conditions and must be accurate for further (3). Therefore, it is important to do the detection of outlier on the data and choose the special method of analysis on data outlier to get the accurate analysis result.

There are two methods to identificate the outlier in this research, namely the use the Scatter Plot and Z score method. Scatter Plot is called the graphic method which is one of general methods used to identificate the outlier with the scatter plot residual observation data using the prediction dependent variable (17). The observational result of outlier by using scatter plot gets some data that are far from the model data in general. This indicates that the data include the outlier. The preliminary research explained that the existence of outlier on the graphic method is identificated if there is one or more research data on the scatter plot that are far from the model data (7).

Next is identificating the outlier by using Z score method. This is because the scatter plot does not give the information spesifically about the existence of outlier on the research data. Z score method is an identification method with univariate, which means the identification of outlier existence on the independent and dependent variables. The Z score result shows that some of independent and dependent variables of the research have absolute Z scores of > 3. The outlier criteria on the Z score method are if every variable of research data has Z Score of> 3 (18).

The outlier is an observation that does not follow some of research data, and the place is far from the central data (19). It is more extreme than the other data in the research. The outlier can make a masking effect and identification error on data non-outlier which become the data outlier (swamping*)* (20). If there is an outlier on the data, it will be removal of data that includes an outlier before doing further analysis. This step is taken if the data outlier is obtained from an error, such as the error in entry or in using the equipment that later has the impact on the error of measuring, resulting in extreme research data (3).

Robust regression is a method used as an alternative on the data outlier (5). There are some estimations such as M estimation, S estimation, and MM estimation (8). The robust regression using MM estimation is one of methods that generally used to overcome data outlier (4). MM Estimation (Method of Moment) is the special part of M estimation (Maximum of Likelihood) which has high efficiency and develops to become an estimator with high breakdown (8).

The research shows that regression analysis using OLS does not get the agreement of regression model. It is because the existence of outlier on the data has confirmed the use of the Scatter Plot and Z score methods. The preliminary research explained that the regression analysis by using OLS is not appropriate enough to be used to estimate the regression parameter on the outlier (21). The OLS will produce a big Mean Square Error (MSE) because of it's refraction to outlier, making the interval credibility becomes bigger (4). It can be seen from the research finding that the MSE value on the regression analysis by using OLS is 2,056.321, with $R^2$ adjusted is 0.009.

Mean Square Error (MSE) is the indicator of an error in the estimation of regression that has been produced. If the residual is smaller, the regression model will be better (22). $R^2$ adjusted is a variate value that can be explained in the regression that has been corrected. If the coefficient of regression produced is bigger, the regression model will be better (23). If compared with the regression analysis result by using robust regression on MM estimation, the MSE produced is smaller, which is 86,026. This MSE value is smaller than MSE value in the OLS regression. Moreover, the $R^2$ adjusted produced is bigger than the OLS regression, which is 0.522. The better regression is regression model producing the smallest MSE with the biggest $R^2$ adjusted for every predictor that has been considered (24). The preliminary result research explained that MM estimation will produce smaller MSE than the OLS (25). Moreover, the calculation of coefficient determination of $R^2$ adjusted on the MM estimation is bigger than the OLS, so the robust regression method was better to used (26).

Some of preliminary researches were conducted to get the optimal model by comparing the results based on the M estimation, S estimation, and MM estimation based on the biggest $R^2$ adjusted with the smallest Mean Square Error (MSE) (24). The other researches mentioned that MM estimation is better than the M and S estimations when there is an outlier on the independent and dependent variables. (4).

MM Estimation method is the method that simultantly has a high breakdown and efficiency by integrating the S estimation as the first estimation and then using M estimation on the iteration calculation process (3). MM Estimation is mostly used because its characteristic is robust. The MM Estimation is called robust with the value characteristic of breakdown point up to 50% and the efficiency up to 95% (8). MM Estimator produced the estimation with high breakdownand holds its efficiency.

MM Estimation combines the robust regression methods of S estimation to get two characteristics simultantly (high breakdown point50%) as the first estimation and M estimation (high efficiency 95%) on the calculation process of the iteration (8). MM Estimation has 50% breakdown obtained from the use of tunning constant up to 1.547 (6). It means that the MM estimation can resolve the outlier until 50% from the overall research data. The use of tunning constant will make the estimator become specific and minimize the smallest quadrate of residual (27).

The data condition with the outlier can be found on the DHF cases, as the preliminary researches mentioned that case data of DHF identificated contain the outlier, so it needs a special analysis method to resolve the outlier on the DHF case data (28). The raising in DHF case in a various districts can be caused by some factors, including the climate change, the change of density, inhabitant distribution, the increasing inhabitant mobility, waste management system, and the water supply that is not satisfying as well as the people being less aware of keeping the environment, making it become not a conducive place for the breeding of *Aedes sp*(29).

The climate change influences the DHF like the *Aedes sp* that is sensitive to climate change such as precipitation, temperature, and damp (30). The optimum climate condition indirectly supports the breeding of *Aedes sp* in the environment that later has impact on the increasing DHF cases (31).

Based on the robust regression using MM estimation, the population density and percentage of healthy home show influence on the IR DHF. If there is an increasing population density up to 50 people per km$^2$, it will become an increasing IR DHF up to 0.5/100,000 people. It is the echology conducted in Bondowoso, East Java that explains the population density influence on the number of DHF cases (32). The high population density makes the increasing DHF cases (33). It is because the population density makes transmission of dengue virus is easy to bring by multiple bites of Aedes sp (32). The population density makes the distance between every house closer that makes it easier for the spread of the virus. A short distance makes *Aedes sp* easy to spread the virus from one people to other people (32).

Moreover, the research finding also shows the percentage of healthy home influence on IR DHF. The research finding explains that if there is decrease in percentage of healthy home up to 50%, it will make an increasing IR DHF up to 26.35/100,000 people. This is because the house and physical environment are not good, emerging a bigger risk to the occurence of DHF than the good house and physical environment (34). The house environment with the Aedes sp breeding place has a risk up to 3.8 times to DHF cases than a house without Aedes sp (35).

The other results that have been obtained are that there are no influence of precipitation and % of PHBS. Precipitation is a number of waterfalls on the certain period of time in milimeters/mm (36). The high precipitation generally makes the high DHF cases, but it is also reducing the DHF on the other conditions. That is because the high precipitation can eliminate breeding place of *Aedes sp,* so there is no place to breeding. The frequency of heavy rain makes running of mosquito larva in the saving water, making less the population of *Aedes sp*(37). Moreover, the percentage of PHBS is related with the cleaning container behavior as a place of saving water, so it does not become *Aedes sp* breeding place. The aim is to reduce the occurrence risk of DHF, butthe breeding of *Aedes sp* is still supported by the container outside, which is uncontrollable. Itcauses there is still a risk of dengue virus infection (38).

Therefore, the development of spread and density of *Aedes sp*is because of the less effective control vector system and the weak of health structure in public (39). The central and regional programs did several control efforts to reduce the number of DHF cases, including the eradication of mosquito nests such as draining habit and closing and burying the secondhand that has the potential as the breeding place for *Aedes sp*, but that action cannot push the national number of DHF, especially in avoiding the fatality because of the DHF. It shows that the effort for pro-active prevention is more need than the reactive prevention. In this case, the public and government must be active to participate in the control and prevention of the DHF (30) .

## Acknowledgement

Thank you to BMKG Karang Ploso Malang, East Java for the helps in providing the research data.

## Reference

1. Pramana S, Yordani R, Kurniawan R, Yuniarto B. Dasar-Dasar Statistika dengan Software R (Konsep dan Aplikasi). Kedua. Bogor: iN Media; 2017.
2. Gujarati DN. Basic Econometrics. 4th ed. Ne: McGraw-Hill/Irwin; 2013.
3. Herawati N, Nisa K, Setiawan E. Analisis Ketegaran Regresi Robust Terhadap Letak Pencilan: Studi Perbandingan. Bull Math. 2011;3(01):49–60.
4. Lainun H, Tinungki GM. Perbandingan Penduga M , S , dan MM pada Regresi Linier dalam Menangani Keberadaan Outlier. J Mat Stat dan Komputasi. 2018;15(1):88–96.
5. Montgomery DC, Elizabeth AP. Introduction to linear regression analysis. 2nd ed. Journal of Applied Statistics. New York: John Wiley & Sons Inc; 1992.
6. Shodiqin A, Aini AN, Rubowo MR. Perbanding Dua Metode Regresi Robust Yakni Metode Least Trimmed Squares (LTS) dengan Metode Estimator-MM (Estimasi-MM) (Studi Kasus Data Ujian Tulis Masuk Terhadap Hasil IPK Mahasiswa UPGRIS ). J Ilm Teknosains. 2018;4(1):35–42.
7. Setiarini Z, Listyani E. Analisis Regresi Robust Estimasi-S Menggunakan Pembobot Welsch dan Tukey Bisquare. J Mat. 2017;6(1):48–55.
8. Yohai VJ. High Breakdown-Point and High Efficiency Robust Estimates For Regression. Ann Stat. 1987;15(4):1580–92.
9. Utami I, Rais. Aplikasi Regresi Kuantil pada Kasus DHF di Kota Palu Sulawesi Tengah. J Ilm Mat dan Terap. 2018;15(1):108–17.
10. Irianto K. Epidemiologi Penyakit Menular dan Tidak Menular : Panduan Klinis. Bandung: Alfabeta; 2014.
11. Masriadi. Epidemiologi Penyakit Menular. Jakarta: Rajawali Pers; 2017.
12. Kementerian Kesehatan RI. Data dan Informasi - Profil Kesehatan Indonesia 2017. 2018. p. 1–184.
13. Kementerian Kesehatan RI. Profil Kesehatan Indonesia. Kementerian Kesehatan RI,. 2016. p. 1–382.
14. Dinas Kesehatan Provinsi Jatim. Profil Kesehatan Provinsi Jawa Timur Tahun 2017. Surabaya; 2017.
15. BPS Provinsi Jawa Timur. Jawa Timur Province in Figures 2018. Surabaya; 2018.
16. Sunaryo S. Mengatasi Masalah Multikolinearitas dan Outlier dengan Pendekatan ROBPCA (Studi Kasus Analisis Regresi Angka Kematian Bayi di Jawa Timur). J Mat Sains dan Teknol. 2011;12(1):1–10.
17. Hidayatulloh FP, Yuniarti D, Wahyuningsih S. Regresi Robust Dengan Metode Estimasi-S Robust Regression Method To Estimate - S. Eksponensial. 2015;6(2):163–70.
18. Tripathy SS. Comparison of Statistical Methods for Outlier Detection in Proficiency Testing Data on Analysis of Lead in Aqueous Solution. Am J Theor Appl Stat. 2013;2(6):233.
19. Barnett V, Lewis T. Outlier in Statistical Data. Third. Boston: John Wiley & Sons Inc, Kluwer Academic Publishers; 1992.
20. Ali S. Hadi, Imon AHMR, Werner M. Detection of Outliers. WIREs Comput Stat. 2009;1(1):57–70.
21. Susanti Y, Pratiwi H, Sulistijowati S. Optimasi Model Regresi Robust Untuk Memprediksi Produksi Kedelai Di Indonesia. In: Seminar Nasional Matematika dan Pendidikan Matematika. 2013. p. 978–9.
22. Yuniastari NLAK, Wirawan IWW. Peramalan Permintaan Produk Perak Menggunakan Metode Simple Moving Average Dan Exponential Smoothing. Sist dan Inform. 2014;9(1):97–106.
23. Dahlan MS. Regresi Linear Disertai Praktik dengan SPSS. Jakarta: Epidemiologi Indonesia; 2012. 1–11 p.

24. Rahman MB, Widodo E. Perbandingan Metode Regresi Robust Estimasi Least Trimmed Square , Estimasi Scale , dan Estimasi Method Of Moment. In: Prosiding Seminar Nasional Matematika (PRISMA). 2018. p. 426–33.
25. Muliyani, Noeryanti. Analisis Regresi Robust Penduga Method of Moment (MM) untuk Mengatasi Data yang Terindentifikasi Pencilan Berdasarkan Data Produksi Kedelai di Indonesia. J Stat Ind dan Komputasi. 2017;2(2):126–35.
26. Nurdin N, Islamiyati A. Penggunaan Regresi Robust Pada Data Yang Mengandung Pencilan Dengan Metode Momen. Mat Stat dan Komputasi. 2014;10(2):114–23.
27. Pradewi ED, Sudarno. Kajian Estimasi-M IRLS Menggunakan Fungsi Pembobot Huber dan Bisquare Tukey pada Data Ketahanan Pangan di Jawa Tengah. Media Stat. 2012;5(1):1–10.
28. Utami ES, Karim A. Perbandingan Regresi Robust dengan OLS pada Produksi Ubi Jalar Provinsi Jawa Tengah Tahun 2015. In: Seminar Nasional Pendidikan, Sains dan Teknologi Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Muhammadiyah Semarang. 2015. p. 83–7.
29. Kementerian Kesehatan RI. Situasi DHF di Indonesia. Kementerian Kesehatan Republik indonesia. 2016. p. 1–12.
30. Ariati J, Anwar A. Model Prediksi Kejadian Demam Berdarah Dengue (DHF) Berdasarkan Faktor Iklim Di Kota Bogor, Jawa Barat. Bul Penelit Kesehat [Internet]. 2014;42(4 Des):249–56. Available from: http://ejournal.litbang.depkes.go.id/index.php/BPK/article/view/3663
31. Rahayu D, Winahju WS, Mukarromah A. Pemodelan Pengaruh Iklim Terhadap Angka Kejadian DHF di Surabaya. J Sains dan Seni ITS. 2012;1:69–74.
32. Prasetyowati I. Kepadatan Penduduk dan Insidens Rate Demam Berdarah Dengue (DHF) Kabupaten Bondowoso, Jawa Timur. Indones J Heal Sci. 2015;5(2):1–12.
33. Ali K, Ma'rufi I. Study of Factors Caused Dengue Haemorrhagic Fever Case Study : Pasuruan , Jawa Timur- Indonesia. J Med Bioeng. 2016;5(2):108–12.
34. Umaya R, Faisya AF, Sunarsih E. Hubungan Karakteristik Pejamu, Lingkungan Fisik Dan Pelayanan Kesehatan Dengan Kejadian Demam Berdarah Dengue (DHF) Di Wilayah Kerja Puskesmas Talang Ubi Pendopo Tahun 2012. J Ilmu Kesehat Masy. 2013;4(November):262–9.
35. Sofia, Suhartono, Wahyuningsih NE. Hubungan Kondisi Lingkungan Rumah dan Perilaku Keluarga dengan Kejadian Demam Berdarah Dengue Di Kabupaten Aceh Besar The Relationship of Home Environmental Conditions and Family Behavior with Genesis Dengue In Aceh Besar. Kesehat Lingkung Indones. 2014;13(1):30–7.
36. Sihombing CG, Nugraheni E, Sudarsono W. the Relationship Between Rainfall, Air Temperature and Wind Speed Effects Dengue Hemorrhagic Fever Case in Bengkulu City At 2009-2014. J Kedokt Diponegoro. 2018;7(1):366–80.
37. Suhermanto, Tunggul STB, Widartono BS. Spatial Analysis on Vulnerability to Dengue Hemorrhagic Fever in Kotabaru Subdistrict , Jambi Municipality , Jambi Province. Trop Med J. 2012;02(1):45–56.
38. Hastuti NM, Dharmawan R, Indarto D. Sanitation-Related Behavior , Container Index , and Their Associations with Dengue Hemorrhagic Fever Incidence in Karanganyar , Central Java. 2017;2:174–85.
39. Kementerian Kesehatan RI. Buletin Jendela Epidemiologi Topik Utama Demam Berdarah Dengue. 2010. p. 1–43.

# Robust Regression Estimation Mm Bisquare On The Factor of DHF In East Java

ORIGINALITY REPORT

| 12% | 11% | 5% | 3% |
|---|---|---|---|
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | |
|---|---|---|
| 1 | www.journal.atmph-specialissues.org<br>Internet Source | 3% |
| 2 | Submitted to Konsorsium Turnitin Relawan Jurnal Indonesia<br>Student Paper | 2% |
| 3 | repo.poltekkes-palangkaraya.ac.id<br>Internet Source | 2% |
| 4 | jurnal.ugm.ac.id<br>Internet Source | 1% |
| 5 | www.researchgate.net<br>Internet Source | 1% |
| 6 | www.scilit.net<br>Internet Source | <1% |
| 7 | eprints.undip.ac.id<br>Internet Source | <1% |
| 8 | garuda.kemdikbud.go.id<br>Internet Source | <1% |

Submitted to Universiti Teknologi MARA

| | | |
|---|---|---|
| 9 | Student Paper | <1 % |
| 10 | Submitted to Higher Education Commission Pakistan<br>Student Paper | <1 % |
| 11 | ijsoc.goacademica.com<br>Internet Source | <1 % |
| 12 | scitcentral.com<br>Internet Source | <1 % |
| 13 | www.readkong.com<br>Internet Source | <1 % |
| 14 | "1st Annual Conference of Midwifery", Walter de Gruyter GmbH, 2020<br>Publication | <1 % |
| 15 | 123dok.com<br>Internet Source | <1 % |
| 16 | journal.unhas.ac.id<br>Internet Source | <1 % |
| 17 | www.repository.cam.ac.uk<br>Internet Source | <1 % |
| 18 | "Proceedings of the Andalas International Public Health Conference 2017", BMC Public Health, 2017<br>Publication | <1 % |

| | | |
|---|---|---|
| 19 | M Irfan, F Virgo, M Y N Khakim, M Ariani, A Sulaiman, I Iskandar. "The dynamics of rainfall and temperature on peatland in South Sumatra during the 2019 extreme dry season", Journal of Physics: Conference Series, 2021<br>Publication | <1 % |
| 20 | M. Imron, Wika Dianita Utami, Hani Khaulasari, Firman Armunanto. "ARIMA MODEL OF OUTLIER DETECTION FOR FORECASTING CONSUMER PRICE INDEX (CPI)", BAREKENG: Jurnal Ilmu Matematika dan Terapan, 2022<br>Publication | <1 % |
| 21 | apfjournal.or.id<br>Internet Source | <1 % |
| 22 | jurnalmka.fk.unand.ac.id<br>Internet Source | <1 % |
| 23 | media.neliti.com<br>Internet Source | <1 % |

| | | | |
|---|---|---|---|
| Exclude quotes | On | Exclude matches | Off |
| Exclude bibliography | On | | |