the 5th ICTROPS
International Conference for Tropical Studies and Its Applications
Islamic Development Bank 4in1 Project
Project Implementation Unit
University of Mulawarman
IsDB Islamic Development Bank
Co-Organized by :

Association Analysis of the Number of Dengue Hemorrhagic Fever Cases in East Kalimantan with the Factors of Geographic, Demographic, and Health Using Spearman Rank Correlation

M. Fathurahman 1*, Ika Purnamasari 2, Surya Prangga 3.
Department of Mathematics, Mulawarman University, Indonesia
Corresponding Author: fathur@fmipa.unmul.ac.id*

Abstract

Dengue Hemorrhagic Fever (DHF) is a mosquito-borne infection caused by the dengue virus, primarily found in the tropics and sub-tropics. The Indonesian country is a tropical climate. During the last three years (2017-2019), there has been an increase in the number of cases of DHF in East Kalimantan Province, Indonesia. This study aims to determine the corresponding factors with the number of DHF cases in East Kalimantan Province, Indonesia, in 2019 by using the Spearman rank correlation method. The result shows that the factors that significantly associated the number of DHF cases in East Kalimantan were the demographic and health factors.
Keywords: DHF, Association Analysis, Spearman Rank Correlation.

1 Introduction

Dengue Hemorrhagic Fever (DHF) is a severe infectious disease and caused by the dengue virus, which is contaminated through the bites of the *Aedes aegypti* and *Aedes albopictus* mosquitoes, which can cause disruption of the capillary blood vessels and the blood clotting system, resulting in bleeding and death when appropriately handled bad. DHF is commonly found in tropical countries such as Indonesia, and it is still a public health problem in Indonesia, especially in East Kalimantan Province. Based on publications from the Ministry of Health (2020), the number of dengue cases in East Kalimantan Province, in 2017 was 2,237 cases with 8 cases of death. Meanwhile, the Incidence Rate (IR) per 100,000 population is 62.57, and the Case Fatality Rate (CFR) is 0.72%. The cases and deaths in 2018 were increased. There are 3,204 and 17, whereas an IR per 100,000 population of 87.81 and CFR of 0.53%. The number of cases dan death in 2019 was 6,723 cases and 44 cases of death, respectively, whereas the IR per 100,000 population value is 180.66, and a CFR value is 0.65%. This IR value is classified as very high. It shows that the number of cases and deaths due to DHF in East Kalimantan in 2019 significantly increased. Therefore, we investigate the factors related to the DHF cases in East Kalimantan, Indonesia, in 2019.

According to Huda, Mukhaiyar, & Pasaribu (2020), the factors associate with the DHF cases were the climate factor. Meanwhile, Sriningsih, Otok, & Sutikno (2021) discusses the association of the DHF cases with the number of health workers, the number of health facilities, the height of an area, and the density of settlements. Following the factors in the previous studies, this study aims to obtain the factors that have associations with the number of DHF cases in East Kalimantan using the Spearman Rank Correlation (SRC) method. The SRC method is a nonparametric correlation technique that can measure the strength of the relationship between two variables, which have the association form is the monotonic function or the nonlinear function (Croux & Dehon, 2010; Hauke & Kossowski, 2011). Research that discusses DHF cases in East Kalimantan using the SRC method is still scarce and limited. Due to the limited data available, the factors thought to be associated with the number of DHF cases in this study were limited to geographic, demographic, and health factors. Geographical factors are represented by the area and altitude of the area. Meanwhile, demographic and health factors are represented by population density and health workers, respectively. Based on data obtained from the East Kalimantan Provincial Health Office (2020) and the Central Statistics Agency of East Kalimantan Province (2020), the pattern of the relationship between the data on the number of DHF cases and each data on total area, altitude, population density, and the number of health workers tends to be in a monotonic function.

2 Methodology

2.1 Research Data

The secondary data was used in this research. The data was got from the Central Bureau of Statistics of East Kalimantan Province (2020). The research unit was ten districts/cities in East Kalimantan Province, Indonesia, in 2019. The data obtained are the number of DHF cases (U), the total area (V_1), the area altitude (V_2), population density (V_3), and the number of health workers (V_4).

2.2 Data Analysis

The steps of data analysis are as follows:

1. Descriptive analysis
   This step aims to determine the characteristics of data, that is presented in the form of tables, graphs, or diagrams.
2. Association analysis of the number of DHF cases was carried out with the following steps:
   a. Identifying the relationship pattern for each pair of variables, namely U and V_1; U and V_2; U and V_3; and U and V_4 using a scatter plot.
   b. Calculate the value of the Spearman correlation coefficient for each pair of variables U and V_1; U and V_2; U and V_3; and U and V_4.
   c. Perform hypothesis testing for each pair of variables U and V_1; U and V_2; U and V_3; and U and V_4, which aims to obtain variables that are significantly related to U.
   d. Interpret each relationship between U and V_1, V_2, V_3, and V_4.
   e. Make conclusions.

2.3 Spearman Rank Correlation Method

Spearman Rank Correlation (SRC) method is a statistical measure of the strength of *monotonicity* relationship between two variables (Kloke & McKean, 2015). Charles Edward Spearman founded the SRC method in 1904 (Conover, 1999). The data form of the SRC method may consist of a bivariate random sample of size n, (U_1, V_1), (U_2, V_2),…, (U_n, V_n). Suppose

$R(U_i)$ is the rank of U_i as compared with the other U values, for $i = 1,2, \dots, n$. That is, $R(U_i) = 1$ if U_i is the smallest of $U_1, U_2, \dots, U_n$, $R(U_i) = 2$ if U_i is the second smallest, and so on, with rank n being assigned to the largest of the U_i. Similarly, let $R(V_i)$ equal 1,2, ..., or n, depending on the relative magnitude of V_i as compared with $V_1, V_2, \dots, V_n$, for each i.

According to Conover (1999), the coefficient correlation value of the SRC method can be obtained by

$$r_S^* = \frac{\sum_{i=1}^{n} R(U_i)R(V_i) - n\left(\frac{n+1}{2}\right)^2}{\left(\sum_{i=1}^{n} R(U_i)^2 - n\left(\frac{n+1}{2}\right)^2\right)^{\frac{1}{2}}\left(\sum_{i=1}^{n} R(V_i)^2 - n\left(\frac{n+1}{2}\right)^2\right)^{\frac{1}{2}}}. \quad (1)$$

If there are no ties, an equivalent but computationally easier form is given by

$$r_S^* = 1 - \frac{\sum_{i=1}^{n}\left(R(U_i) - R(V_i)\right)^2}{n(n^2-1)} = 1 - \frac{6W}{n(n^2-1)} \quad (2)$$

where W represents the entire sum in the numerator. This form is equivalent only if there are no ties. If there are many ties use Equation (1). If a moderate number of ties is present in the data, Equation (2) is recommended for computational simplicity since the difference between Equations (1) and (2) will be slight.

However, the coefficient correlation of the SRC method has positive or negative values (i.e., $-1 \le r_S^* \le 1$) and the strengh of correlation. The positive and negative values of the SRC method are sometimes called the direct correlation. The interpretation of the direct and strength of the correlation is presented in Table 2.

Table 2. Interpretation of the Correlation Coefficient of the SRC Method

Coefficient Correlation Values	Interpretation
$r_S^* = 1$	Perfect correlation
$0.9 \le r_S^* < 1$	Very high correlation
$0.7 \le r_S^* < 0.9$	High correlation
$0.5 \le r_S^* < 0.7$	Moderate correlation
$0.3 \le r_S^* < 0.5$	Low correlation
$0 < r_S^* < 0.3$	Little if any correlation
$r_S^* = 0$	No correlation

The SRC method is often used as a test statistic to test for independence between two random variables. The test statistic is given by Equation (1) or Equation (2). Exact quantiles of r_S^* when U and V are independent are given in Spearman's table for $n \le 30$ and no ties. For larger n, or many ties, the qth quantile of r_S^* is given approximated by

$$W_q = \frac{Z_q}{\sqrt{n-1}} \quad (3)$$

where Z_q is the standard normal quantile found in the standard normal distribution table (Conover, 1999). The coefficient correlation of the SRC method is sensitive to some types of dependence, so it is better to be specific as to what type of dependence may be detected. Therefore, the hypotheses take the following form:

1. Two-Tailed Test
   H_0 : The U_i and V_i are mutually independent
   H_1 : Either (a) there is a tendency for the larger values of U to be paired with the larger values of V, or (b) there is a tendency for the smaller values of U to be paired with the larger values of V
   Reject H_0 at the level of significance (α) if the absolute value of r_S^*, $|r_S^*|$, is greater than its $1-\alpha/2$ obtained from the Spearman's table or Equation (3). The approximate two-tailed P-values is
   $$P\text{-value} = 2P\left(Z \geq |r_S^*|\sqrt{n-1}\right) \qquad (4)$$
   using the table of standard normal distribution (Conover, 1999).
2. Lower-Tailed test for Negative Correlation
   H_0 : The U_i and V_i are mutually independent
   H_1 : There is a tendency for the smaller values of U to be paired with the larger values of V, and vice versa
   Reject H_0 at the level α if $r_S^* < -W_{1-\alpha}$, where $W_{1-\alpha}$ is found in Spearman's table or from Equation (3). The approximate lower-tailed P-value is
   $$P\text{-value} = P\left(Z \leq r_S^*\sqrt{n-1}\right) \qquad (5)$$
   using the table of standard normal distribution (Conover, 1999).
3. Upper-Tailed Test for Positive Correlation
   H_0 : The U_i and V_i are mutually independent
   H_1 : There is a tendency for the larger values of U and V to be paired together
   Reject H_0 at the level α if $r_S^* > W_{1-\alpha}$, where $W_{1-\alpha}$ is found in the Spearman's table or Equation (3). The approximate upper-tailed P-value is
   $$P\text{-value} = P\left(Z \geq r_S^*\sqrt{n-1}\right) \qquad (6)$$
   using the table of standard normal distribution (Conover, 1999).

2.4 Dengue Hemorrhagic Fever

Dengue was a mosquito-borne viral contamination found in tropical and sub tropical climates worldwide, primarly urban and semi-urban. The virus responsible for causing dengue was called dengue virus (DENV). There are four DENV serotypes, meaning that it is possible to be infected four times. The clinical manifestations of DENV contamination range from asymptomatic contamination or a mild flu-like syndrome, also referred as dengue fever (DF), to the more severe and life-threatening forms, dengue hemorrhagic fever (DHF) and dengue shock syndrome (WHO, 2011).

DHF was defined by having at least the first two of the following four clinical manifestations of: 1) sudden onset acute fever of 2 to 7 days duration; 2) unconstrained hemorrhagic positive Tourniquet test; 3) hepatomegaly; and 4) circulatory system failure, in combination with hematological criteria of thrombocytopenia ($\leq$ 100,000 cells/mm^3) and an increased hematocrit over 20%. Suspected DHF cases according to those point of reference were further evaluated which DHF cases were grouped into probable and confirmed cases. A probable case was referred as clinically suspect patients espoused by positive dengue serology (positive anti-DENV IgM in critical or convalescent serum sample and a fourfold intensify in IgG between the critical and the convalescent samples) or when an expected DHF patient was linked at the exact location and time to other confirmed DHF cases. A confirmed cases was referred as a case with laboratory confirmation through DENV isolation, or viral antigen or RNA detection in serum. This categorization has continually been used nationwide by hospitals and Community Health Centres in Indonesia. All probable or confirmed DHF cases were reported and included in the surveillance system (Karyanti et al., 2014; Harapan et al., 2019).

3 Results and Discussion

3.1 Analysis of Descriptive Statistics

The descriptive statistics of research data are displayed in Table 3.

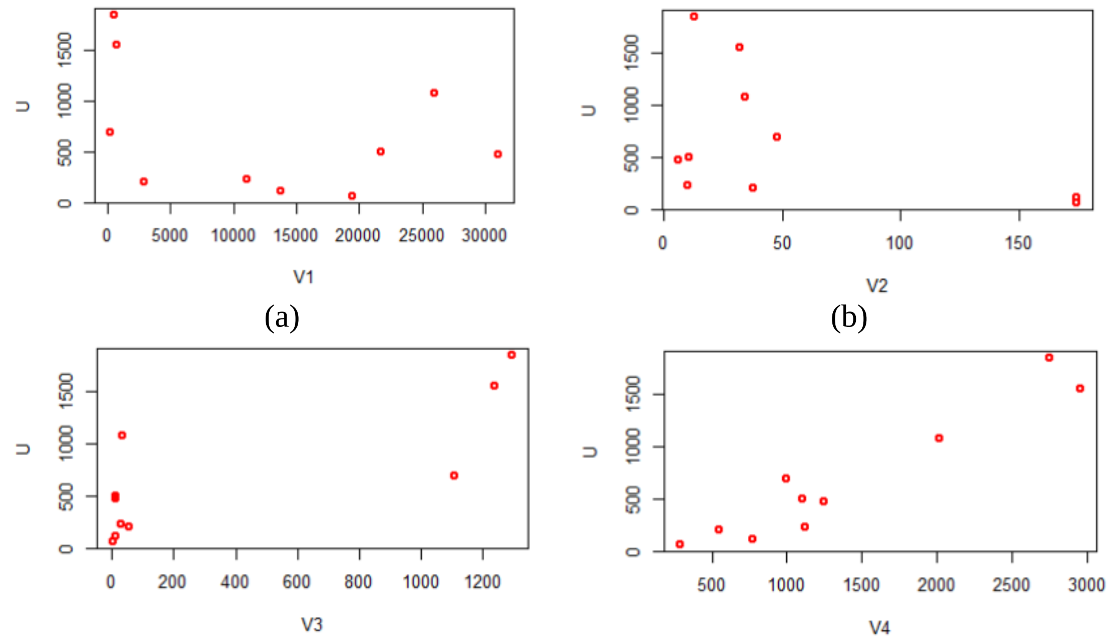
Table 3. Descriptive Statistics of Research Data

Notation	Variables	Minimum	Maximum	Mean	Standard Deviation
U	The number of DHF cases	66	1838	672	122
V_1	The total area	163.1	31051.7	12734.7	11494.05
V_2	The area altitude	5.98	174.63	54.1	65.03897
V_3	Population density	1.36	1297.74	379.65	579.0708
V_4	The number of health workers	287	2960	1381	903

Based on Table 3, the highest number of DHF cases was in Balikpapan City. The lowest was in Mahakam Ulu District. Furthermore, the largest area was Kutai Timur District, and the smallest was Bontang City. The highest area was Mahakam Ulu District, whereas the smallest was Kutai Timur District. The most density of population was Balikpapan City, and the sparsely was Mahakam Ulu District. Meanwhile, most health workers in Samarinda City, and a few were in Mahakam Ulu District.

3.2 Association Analysis for The Number of DHF Cases

The initial step of the association analysis was to identify the relationship pattern using the scatter plots for each pair of variables, namely the number of DHF cases and the total area (Figure 1a); the number of DHF cases and the area altitude (Figure 1b); the number of DHF cases and population density (Figure 1c); and the number of DHF cases and the number of health workers (Figure 1d). Figure 1 shows that the relationship pattern of paired data has a monotonic (nonlinear) form. Therefore, the SRC method was appropriately used.



(a) (b) (c) (d)

Figure 1. The Relationship Pattern of Research Data

Spearman's test statistic performed hypothesis testing for the SRC method. The Spearman's coefficient correlation and Spearman's test statistic values were employed. Based on data analysis using R software with *stats* package, the results of Spearman's coefficient correlation and P-values were presented in Table 4.

Table 4. Spearman's Coefficient Correlation $(\boldsymbol{r}_S^*)$ and Probability (P) Values of Spearman's Test Statistic for the Number of DHF Cases

Variables	$\boldsymbol{r}_S^*$	***P*-Value**
U and V_1	-0.3333	0.3488
U and V_2	-0.4012	0.2505
U and V_3	0.8061	0.0082*
U and V_4	0.8667	0.0027*

*) Significant at the 5% significance level.

Table 4 shows the statistical tests of the two paired variables (i.e., U and V_3 and U and V_4) have the P-values less than the significant level. This results indicate that the number of DHF cases in East Kalimantan was significantly associated with population density and health workers. The strong correlation of the number of DHF cases with population density and health workers was very strong, respectively.

4 Conclusions

During the last three years (2017-2019), there has been an increase in cases and deaths due to DHF in East Kalimantan Province, Indonesia. The relationship pattern of the number of DHF cases in East Kalimantan, Indonesia, in 2019 with the factors of geographic, demographic, and health was a monotonic function. Therefore, It can be analyzed by the SRC method. The factors that have significant association with the number of DHF cases in East Kalimantan, Indonesia, in 2019 were the demographic and health factors (i.e., population density and the number of health workers).

ACKNOWLEDGEMENT

We would like to thank the Islamic Development Bank 4in1 Project, Project Implementation Unit Mulawarman University for funding this work.

REFERENCES

Conover, W. J. (1999). *Practical Nonparametric Statistics* (3rd ed.). New York: John Wiley & Sons.

Croux, C., & Dehon, C. (2010). Influence functions of the Spearman and Kendall Correlation Measures. *Stat. Methods. Appl.*, 497-515.

Harapan, H., Michie, A., Mudatsir, M., Sasmono, R. T., & Imrie, A. (2019). Epidemiology of Dengue Hemorrhagic Fever in Indonesia: Analysis of five decades data from the National Disease Surveillance. *BMC Research Notes, 12*(350), 1-6.

Hauke, J., & Kossowski, T. (2011). Comparison of Values of Pearson's and Spearman's Correlation Coefficients on the Sama Sets of Data. *Quaestiones Geographicae, 30*(2), 87-93.

Huda, N. M., Mukhaiyar, U., & Pasaribu, U. S. (2020). Forecasting Dengue Fever Cases Using Autoregressive Distributed Lag Model with Outlier Factor. *Proceeding of IICMA 2019* (pp. 1-7). Pontianak, Indonesia: AIP Publishing.

Karyanti, M. R., Uiterwaal, C. S., Kusriastuti, R., Hadinegoro, S. R., & Rovers, M. M. (2014). The Changing Incidence of Dengue Haemorrhagic Fever in Indonesia: A 45-year registry-based analysis. *BMC Infectious Diseases*, 1-7.

Kloke, J., & McKean, J. W. (2015). *Nonparametric Statistical Methods Using R.* Boca Raton: CRC Press.

Ministry of Health. (2020). *Health Profile of Indonesia 2019.* Jakarta: Ministry of Health.

Sriningsih, R., Otok, B. W., & Sutikno. (2021). Factor Affecting the Number of Dengue Fever Cases in West Sumatera Province Using the Multivariate Adaptive Regression Splines (MARS) Approach. *Proceeding of ICW-HDDA-X 2020* (pp. 1-6). Jatinangor, Indonesia: IOP Publishing.

The Central Statistics Agency of East Kalimantan Province. (2020). *East Kalimantan Province in Figures 2020.* Samarinda: The Central Statistics Agency of East Kalimantan Province.

The East Kalimantan Provincial Health Office. (2020). *Health Profile of East Kalimantan Province 2019.* Samarinda: The East Kalimantan Provincial Health Office.

WHO. (2011). *Comprehensive Guidelines for Prevention and Control of Dengue and Dengue Haemorrhagic Fever.* New Delhi: WHO Regional Publication SEARO.

COMMITTEE
5th INTERNATIONALCONFERENCEON
TROPICALSTUDIESANDITSAPPLICATION
Gedung *Integrated Laboratory* Universitas Mulawarman Lt.3, Jalan Kuaro,Kampus Gunung Kelua Samarinda –East Kalimantan 75123
E-mail:ictrops@unmul.ac.id| Website:ictrops-idb.unmul.ac.id

Letter of Acceptance
No: 027/Panpel/2021

Dear,

M. Fathurahman, Ika Purnamasari, Surya Prangga
fathur@fmipa.unmul.ac.id

We are pleased to inform you that your abstract, entitled:

Association Analysis of the Number of Dengue Hemorrhagic Fever Cases in East Kalimantan with the Factors of Geographic, Demographic, and Health Using Spearman Rank Correlation

Has been accepted for Oral Presentation at the 5th ICTROPS Conference.
After the acceptance, you may re-submit your abstract, prepare the full-paper and also presentation file or video based on the information bellow:

File	File Name	Deadline	Submission link
Abstract	Subtopic_ID_Abstract_Last Name of Corresponding Author For example:SEE_402_Abstract_Sulaeman	26 September 2021	https://bit.ly/Abstract-Ictrops5
Full-paper	Subtopic_ID_Paper_Name of Corresponding Author For example: SEE_402_Paper_Sulaeman	1 October 2021	https://bit.ly/FullPaper-Ictrops5
Presentation slides or video (duration 5 – 7 minutes)	Subtopic_ID_Name of Corresponding Author For example: SEE_402_Sulaeman	3 October 2021	https://bit.ly/PPT-Ictrops5

The review committee will evaluate your article to ensure that it meets the requirements for the conference.

If you have any question or concern, please do not hesitate to contact us through our email address ictrops@unmul.ac.id. Looking forward to meeting you in our conference.

Yours sincerely,

Nurul FitriyahSulaeman, PhD.
Chairman of 5th ICTROPS Commitee