

# Quadratic Discriminant Analysis of Dengue Viruses Disease Incidence in Palembang

<sup>1</sup>Yulia Resti, <sup>2</sup>Noriszura Ismail and <sup>3</sup>Marieska Verawaty

<sup>1</sup>Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sriwijaya, Jl. Raya Palembang-Prabumulih Km.32 Inderalaya 3066, Sumatera Selatan, Indonesia

<sup>2</sup>Pusat Pengajian Sains Matematik, Fakulti Sains dan Teknologi, Universiti Kebangsaan Malaysia, 43600, UKM Bangi, Selangor Darul Ehsan, Malaysia

<sup>3</sup>Jurusan Biologi, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sriwijaya, Jl. Raya Palembang-Prabumulih Km.32 Inderalaya 3066, Sumatera Selatan, Indonesia

## Article history

Received: 06-12-2016

Revised: 18-05-2017

Accepted: 27-05-2017

## Corresponding Author:

Yulia Resti

Jurusan Matematika, Fakultas

Matematika dan Ilmu

Pengetahuan Alam, Universitas

Sriwijaya, Jl. Raya Palembang-

Prabumulih Km.32 Inderalaya

3066, Sumatera Selatan,

Indonesia

Email: yulia\_resti@mipa.unsri.ac.id

**Abstract:** Palembang is classified as a city that has tropical climate with abundant rainfall, even in the driest month. These characteristics indicate that the city has a high potential for the occurrence of dengue virus disease. The areas of incidence of dengue virus disease in Palembang can be classified into five areas; South, Center, North, West and East. The aim of this paper is to map the incidence of dengue virus disease into these areas based on significant factors using the quadratic discriminant analysis. The stratified sampling technique is applied to obtain the respondents who have family members suffering from dengue disease. The results showed that the incidence of dengue virus disease in each area is significantly affected by all factors, except the age of patient, hygienic behavior of respondent, age and occupation of family head and number of bedrooms. The results of the Wilk's lambda test and the F test indicate that gender and blood type of patient; age, education, occupation, income and knowledge of respondent; occupation and income of family head; type of residence; number of bathrooms that have tubs and clean water source are relatively different in each area. The overall correct percentage of the mapping result is 66.7%.

**Keywords:** Dengue Virus Disease Incidence, Discriminant Analysis, Palembang

## Introduction

Palembang is the capital of South Sumatra province which has an area of 400.61 km<sup>2</sup> (CBS, 2014). The region is divided by the Musi river into two areas namely Seberang Ilir (245.97 km<sup>2</sup>) and Seberang Ulu (154.64 km<sup>2</sup>). In topography, the two regions are different, Seberang Ulu generally has a relatively flat topography, while Seberang Ilir has a topography that varies between 4-20 m above sea level. Palembang has a tropical climate with a relative humidity wind, the temperature of 23.4-31.7°C and the average rainfall of 227.23 mm per year (BMKG, 2008). The tropical climate conditions of Palembang has the potential of breeding the *Aedes Aegypti* mosquito which is the cause of dengue virus disease.

According to the Indonesian Ministry of Health 2011 in Ismah (2014), in addition to the tropical climate, another factor that has the potential of breeding the *Aedes Aegypti* mosquito are the economic condition,

type of housing, population densities, hygienic behavior of population and health management. Salamah *et al.* (2012) explored the effects of climatic and socio-economics factors towards the number of dengue virus disease in 13 districts of East Java. They employed the semi parametric panel regression approach and compared this approach with the standard panel regression. Their results showed that the number of dengue virus disease incidences is significantly influenced by the per-capita income and the number of people below 15 years. They also noted that the dengue virus disease incidence in East Java is optimum under the humidity of 82-87%, the temperature of 22-27°C and the rainfall of 1500 mm 3670 mm, with Surabaya as the most responsive district with respect to the change of climate variables. Hii (2013) revealed that the dengue virus disease incidence in Singapore (2000-2012) is significantly affected by the mean temperature and the cumulative rainfall. Toan (2015)

identified the factors of dengue virus diseases incidence in Hanoi based on the demographic characteristics of patients. She employed the logistic regression and found that living near the open sewers increased the risk of incidence at 7.9 times, which is the highest risk compared to other associated factors.

The spreading area of dengue virus diseases incidence in Palembang can be divided into five areas; Center, North, West, East in the Seberang Ilir and South in Seberang Ulu. Ismah (2014) showed that in years 2009-2013, the distribution of dengue virus disease incidence in Palembang was mostly located in the North and Central areas, where both areas have a high population density, a high percentage of healthy home, a low percentage of poor people and a low Larvae Free Index (LFI). The survey results showed that the majority of population in both areas have a lot of water reservoirs.

In this study, we explore the associated factors that influence the dengue virus diseases incidence in every area in Palembang using the quadratic discriminant analysis. In the economic and financial research, the application of discriminant analysis has been reviewed by Siqueira *et al.* (2012) who selected and evaluated the discriminate factors between the most profitable and less profitable stocks. They investigated the fundamental and the financial variables and the beta coefficient of CAPM of stocks traded on Sao Paulo Stock Exchange (BM and FBovespa) in 2006-2010. Li and Shao (2015) also proposed the discriminant analysis, especially quadratic, for the sparse high dimensional data to classify the tumor colon tissues. Arasi *et al.* (2016) applied the discriminant analysis to analyze the factors that influence the success of web for online business service. Their research found that the factors of availability and compliance in e-service were significantly different.

## Materials and Methods

### Data Collection and Description

We collected the characteristics of dengue virus diseases incidence for the five areas in Palembang using a survey data. The survey population comprised of housewives who had family members suffering from dengue virus disease in year 2015. According to PKSS (2015), the number of dengue virus disease incidence in Palembang in 2015 was 979 people. The respondents has lived for at least one year in these areas. The stratified sampling technique was used to find the respondents from each area, where the number of samples is determined by the Slovin formula. The percentage of incidence of dengue virus disease that spread in each area is presented in Table 1.

Table 1. Proportion of dengue virus disease incidence for each area

Area	Sample (%)
South	15.09
Central	20.70
North	20.00
West	9.470
East	34.74

The questionnaire of the survey consisted of five parts. The first part consisted of information on the family members suffering from dengue virus disease, such as age, gender and blood group of patient. The second part consisted of information on the respondent and her husband, such as age, educational level, occupation and income per month. The third part explored the respondent's knowledge on dengue virus disease which was assessed based on 21 questions. The answers' options consist of 'true' or 'false', whereas the questions were focused on dengue virus disease transmission and vector and clinical manifestations. The fourth part explored the behavior and practices of prevention towards the incidence of dengue virus disease, such as the use of mosquito repellent, mosquito nets, mosquito wiring and fogging; the management of used containers, the keeping of fish-eating mosquito larvae and the cropping of mosquito repellent. This behavior was assessed based on 17 questions, where the answer options consist of 'always', 'sometimes' or 'never'. The fifth part consisted of information on types of residence, number of bedrooms, number of bathrooms that have tubs and clean water source.

### Quadratic Discriminant Analysis

Consider  $X = (X_1, X_2, \dots, X_p, \dots, X_{17})^T$  as the vector of independent variable or characteristic for the incidence of dengue virus disease. Let  $Y_i$  be the  $i$ -th area of dengue virus disease incidence,  $i = 1, 2, \dots, j, \dots, k$  and each of  $Y_i$  has  $n_i$  sample, so that the total number of samples is

$$n = \sum_{i=1}^k n_i. \text{ Let } S_i \text{ be the covariance matrix and } p_i \text{ be the}$$

prior probability of the  $i$ -th area respectively. The purpose of quadratic discriminant analysis in this study is to map  $n$  samples into 5 areas based on the quadratic discriminant score  $Z_i$ . A respondent who has a family member suffering from dengue virus disease is allocated to area  $Y_j$  if it has the largest quadratic discriminant score compared to other areas (the smallest quadratic distance). The quadratic discriminant score of the  $i$ -th area is defined as:

$$Z_i = -\frac{1}{2} \ln |S_i| - \frac{1}{2} X^T S_i^{-1} X + \bar{X}_i^T S_i^{-1} X - \frac{1}{2} \bar{X}_i^T S_i^{-1} \bar{X}_i + \ln p_i \quad (1)$$

This score function was under the assumption of unequal multivariate normal distributions among groups.

In this study, we considered 17 categorical variables in mapping the  $n$  samples into the 5 areas, namely age ( $X_1$ ), gender ( $X_2$ ) and blood group ( $X_3$ ) of the patient; age ( $X_4$ ), education ( $X_5$ ), occupation ( $X_6$ ), income per month ( $X_7$ ), knowledge of dengue virus disease ( $X_8$ ) and behavior about practices of prevention towards the incidence of dengue virus disease ( $X_9$ ) of the respondent; age ( $X_{10}$ ), education ( $X_{11}$ ), occupation ( $X_{12}$ ) and income per month ( $X_{13}$ ) of the head of family; type of residence ( $X_{14}$ ); number of bedrooms ( $X_{15}$ ), number of bathrooms that have tubs ( $X_{16}$ ) and clean water source ( $X_{17}$ ). Variables  $X_2$  and  $X_8$  have 2 categories, variables  $X_1$  and  $X_9$  have 3 categories, variable  $X_3$  has 4 categories, while other variables have 5 categories. The categorical variables were transformed into continuous variable using the Successive Interval method.

## Result and Discussion

We employed several tests to explore the assumptions of quadratic discriminant score. The result of Box-M test ( $\alpha = 5\%$ ) showed that the covariance matrices of North, South, Center, West and East areas were not homogen (Box-M test = 1802.15 >  $F_{(\alpha, df)} = 670.66$ ). The result of the test of equality of means for the areas using the Wilk's lambda and Fas given in Table 2 showed that gender and blood type of family member who suffered from dengue ( $X_2$  and  $X_3$ ), age, education, occupation, income and knowledge of respondent ( $X_4$ ,  $X_5$ ,  $X_6$ ,  $X_7$  and  $X_8$ ), occupation and income of family head ( $X_{11}$  and  $X_{13}$ ), type of residence, number of bathrooms that have tubs and clean water source ( $X_{14}$ ,  $X_{16}$  and  $X_{17}$ ) were relatively different in each area (for Wilk's lambda test, its value is close to one and for F test, p-value <  $\alpha$ ).

The dengue virus diseases incidence in the North, South, Central, West or East areas were affected by all factors, except the patient's age ( $X_1$ ), the behavior of practices of prevention towards the incidence of dengue virus disease of the respondent ( $X_9$ ), the age and occupation of the family head ( $X_{10}$  and  $X_{12}$ ) and the number of bedrooms ( $X_{15}$ ).

The results of the Box-M test and the equality of means test were also supported by the plot Mahalanobis Distance with Chi-Square as shown in Fig. 1. The plots were not a straight line, indicating that the covariance matrix is not homogeneous and the mapping of incidence of dengue virus can be constructed using different covariance matrix as shown in Equation 1.

The quadratic discriminant score for the South, Center, North, West and East areas, together with the largest score in each area are shown in Table 3.

Table 4 provided the successive number of respondents which were mapped into the 5 areas (South, Central, North, West and East) based on the largest quadratic discriminant score,  $Z_i$ . The results showed that as many as 30.2, 66.1, 66.7, 44.4 and 88.0% of respondents were successively mapped into each area, so that the overall correct percentage of the mapping is 66.7%. For others, as many as 69.8, 33.9, 33.3, 55.6 and 11.1% of respondents were successively mapped to other areas. The results showed that this model is sufficient for mapping the incidence of dengue fever into five areas in Palembang based on the correct percentage of respondents mapping.

One of the main benefits of mapping the incidence of dengue virus disease is that the results can provide the characteristics that significantly affect the incidence of dengue in each area and the information can be used by the government in the management of prevention and control of disease.

Table 2. Tests of equality of means

	Independent variable	Wilk's lambda	F	p-value
Family member who get DVD	$X_1$ , age	0.99	0.76	0.55
	$X_2$ , gender	0.62	43.90	0.00
	$X_3$ , blood type	0.91	6.70	0.00
Respondent	$X_4$ , age	0.92	6.16	0.00
	$X_5$ , education	0.91	6.86	0.00
	$X_6$ , occupation	0.91	7.30	0.00
	$X_7$ , income per month	0.86	11.47	0.00
	$X_8$ , knowledge	0.79	18.91	0.00
	$X_9$ , behavior	0.99	0.21	0.94
Head of family	$X_{10}$ , age	0.99	0.95	0.44
	$X_{11}$ , education	0.78	20.29	0.00
	$X_{12}$ , occupation	0.99	1.08	0.37
	$X_{13}$ , income per month	0.91	6.79	0.00
Facility of residence	$X_{14}$ , type of residence	0.89	8.60	0.00
	$X_{15}$ , number of bedrooms	0.97	2.44	0.05
	$X_{16}$ , of bathrooms that have the tub	0.87	10.35	0.00
	$X_{17}$ , clean water source	0.92	6.28	0.00

Table 3. The largest quadratic discriminant score

Respondent	Quadratic discriminant score					The largest $Z_i$
	$Z_1$	$Z_2$	$Z_3$	$Z_4$	$Z_5$	
1	180,02	-1357,13	270,75	-14,99	294,86	$Z_5$
2	308,64	2165,74	613,17	-31,94	102,07	$Z_2$
3	-136,07	-1363,73	-243,40	101,58	125,05	$Z_5$
4	-55,80	1064,99	67,49	-241,54	149,34	$Z_2$
5	322,65	591,69	-679,09	-28,45	-160,14	$Z_2$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
285	-210,94	2279,19	64,80	-196,78	-109,43	$Z_2$

Table 4. Mapping result

Observed	Mapped					Percent Correct
	South	Center	North	West	East	
South	13	8	12	4	6	30.2%
Center	8	39	8	1	3	66.1%
North	6	8	38	2	3	66.7%
West	4	3	1	12	7	44.4%
East	4	1	4	2	88	88.9%
Overall percentage						66.7%

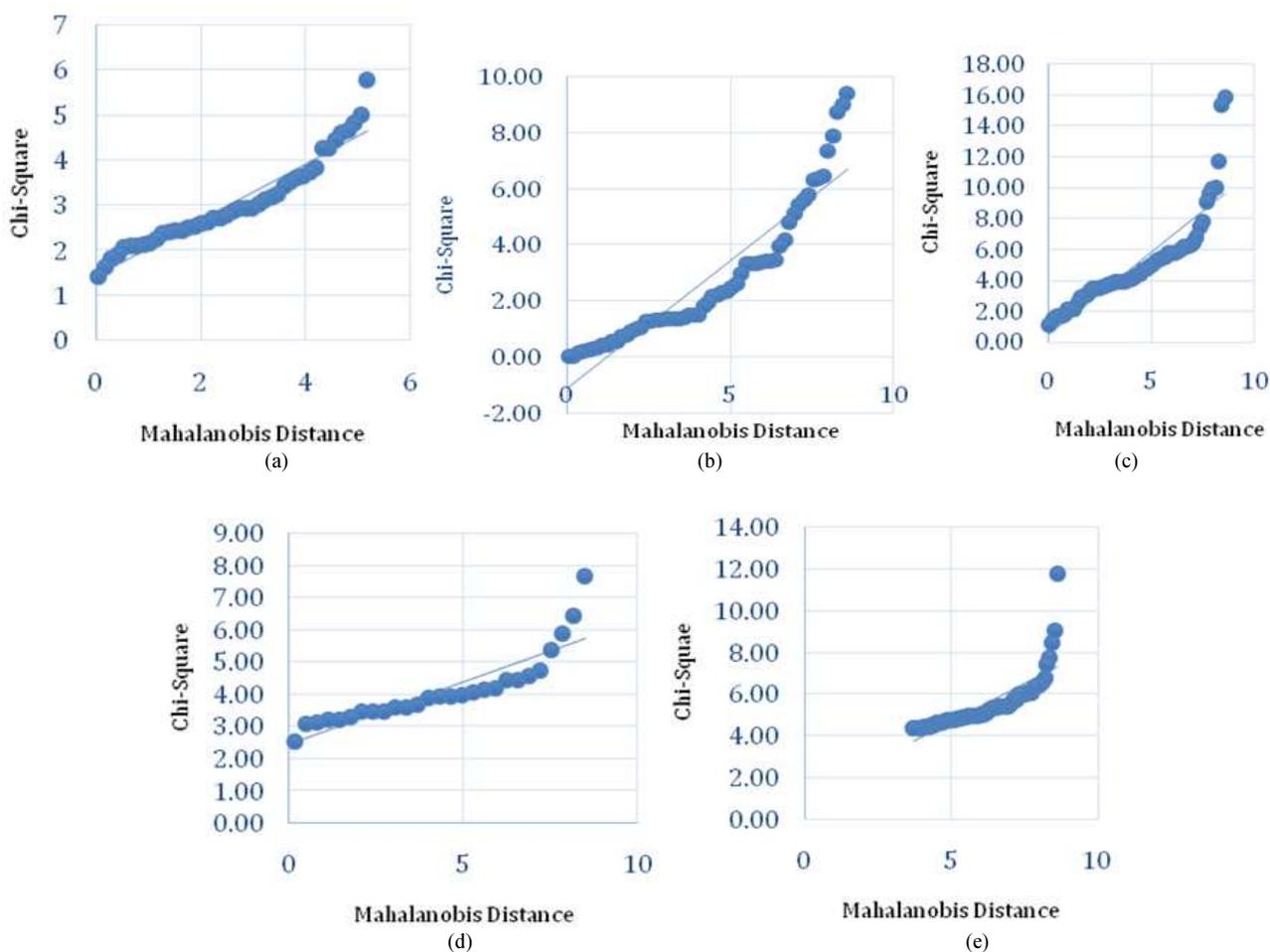


Fig. 1. Plot Mahalanobis Distance for (a) South, (b) Center, (c) North, (d) West, (e) East

## Conclusion

This paper has applied the quadratic discriminant analysis approach in mapping the dengue viruses disease into 5 areas of Palembang; South, Central, North, West and East. The results showed that the mapping of dengue virus disease incidence in each area is significantly affected by all factors, except age of family member suffering from dengue viruses disease, respondent's hygienic behavior, age and occupation of family head and number of bedrooms. The gender and blood group of patient; age, education, occupation, income and knowledge of respondent; occupation and income of family head; type of residence, number of bathrooms that have tubs; and clean water source are relatively different in each area based on the results of the Wilk's lambda test and the F test. The overall correct percentage of the mapping results is 66.7%. This percentage, which is above 50%, indicates that the proposed model is sufficient for mapping the incidence of dengue fever into five areas in Palembang.

## Acknowledgement

The authors are very grateful to Univeritas Sriwijaya, which funded this research (International Research Collaboration Schem, Anggaran DIPA Universitas Sriwijaya, No.042.01.2.400953/2016). This paper is a part of the research. The author are grateful to the Universiti Kebangsaan Malaysia and also many thanks to reviewers for providing valuable comments that help to improve this manuscript.

## Author's Contributions

All authors contributed extensively to the work presented in this paper.

**Yulia Resti:** Built the concept of survey contents, collected and analyzed the data, wrote a draft of the article and revised it for intellectual contents.

**Noriszura Ismail:** Proofreading, gave the technical support and conceptual suggestions and edited the manuscript.

**Marieska Verawaty:** Built the questionnaire for the survey and wrote the supplementary information.

## Ethics

The eligible housewife as respondent in this research will receive an explanatory statement about the research at first, if they are willing (informally) to provide information related to this study, then the survey will be continued. Their identity will remain anonymous in the publications resulting from this study.

## References

- Arasi, F.E.M., S. Govindarajan and A. Subbarayan, 2016. Discriminant analysis of web services success ability. *J. Applied Sci.*, DOI: 10.3923/jas.2016.223.229
- BMKG, 2008. Laporan iklim kota Palembang.
- CBS, 2014. Central Bureau of Statistics.
- PKSS, 2015. Jumlah Kasus DBD Menurut Jenis Kelamin, Kecamatan, dan Puskesmas Kota Palembang Tahun. Seksi Pengendalian and Pemberantasan Penyakit, Dinas Kesehatan Provinsi, Profil Kesehatan Sumatera Selatan.
- Hii, Y.L., 2013. Climate and dengue fever: Early warning based on temperature and rainfall Dissertation, Department of Public Health and Clinical Medicine Epidemiology and Global Health. Umea University, Sweden.
- Ismah, 2014. Distribusi spasiotemporal penyakit demam berdarah dengue di kota Palembang, Provinsi Sumatera Selatan, 2009-2013, Skripsi Prodi Kesehatan Masyarakat, Fakultas Kedokteran dan Ilmu Kesehatan, Universitas Islam Negeri Syarif Hidayatullah Jakarta.
- Li, Q. and J. Shao, 2015. Sparse quadratic discriminant analysis for high dimensional data. *Stat. Sinica*, 25: 457-473. DOI: 10.5705/ss.2013.150
- Salamah, M.K., N. Yussanti and H. Kuswanto, 2012. On the influence of climate and socio-economic condition to the dengue incidences: A semiparametric panel regression approach. *Am. J. Environ. Sci.*, 8: 661-667. DOI: 10.3844/ajessp.2012.661.667
- Siqueira, E., T. Otuki and N. da Costra, 2012. Stock return and fundamental variables: A discriminant analysis approach. *Applied Math. Sci.*, 6: 5719-5733.
- Toan, D.T.T., 2015. Climate change and dengue transmission in Vietnam: An integrated assessment. PhD Thesis, Maastricht University.