The Mapping of Incidence of Dengue Fever in Palembang: A Multinomial Regression Aprroach

By Marieska Verawaty

International Journal of Applied Mathematics and Statistics,
Int. J. Appl. Math. Stat.; Vol. 56; Issue No. 4; Year 2017,
ISSN 0973-1377 (Print), ISSN 0973-7545 (Online)

The Mapping of Incidence of Dengue Fever in Palembang: A Multinomial Regression Aprroach

Yulia Resti[1], Noriszura Ismail[2], Marieska Verawaty[3]

[1]Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Sriwijaya
email : yulia_resti@mipa.unsri.ac.id

Jl. Raya Palembang-Prabumulih Km.32 Inderalaya 3066, Sumatera Selatan, Indonesia
[2]Pusat Pengajian Sains Matematik, Fakulti Sains dan Teknologi,
Universiti Kebangsaan Malaysia
43600, UKM Bangi, Selangor Darul Ehsan, Malaysia

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

ABSTRACT

Palembang, the capital of South Sumatra province, has a high potential for the occurrence of dengue fever. The purpose of this paper is to map the incidence of dengue fever into five areas in Palembang based on the multinomial regression approach. The respondent of this research are housewife who has a family member experiencing dengue fever. The results show that the decision to map the incidence of dengue fever for respondents is affected by family member who get dengue fever (age, gender and blood type), respondent (age, education level, and knowledge of dengue fever), and head of family (age). This model is good enough for mapping the incidence of dengue fever into five areas in Palembang based on both the Nagelkerke value (0.784) and the correct percentage of respondents mapping (65.6%).

Keywords: The incidence of dengue fever, multinomial logistic regression, Nagelkerke.

Mathematics Subject Classification: 62J12, 62G99

Computing Classification System: I.4

1. INTRODUCTION

To date, the desease caused by the flaviviruses or known as dengue virus is still a major health concern worldwide. The disease and its complications cause an estimated 50 to 100 million infections, a half-million hospitalizations, and 22,000 deaths each year in the world. In the last 50 years, its incidence has increased 30-fold (Knowlton et.al., 2009). In the beginning of 2016, South Sumatra, Indonesia, which already has a number of the dengue cases as many as 873 cases has been determined by the Minister of Health of the Republic of Indonesia (MHRI) as a province having a case of extraordinary event of the dengue. Based on data from the Provincial Health Office of South Sumatra, there were 3,549 dengue fever cases during the year 2016, while in 2015 there are 3,401 cases and in 2014 there were only 1,612 cases.

In Indonesia, the dengue studies based on a survey have been done some researchers. Sari (2012) who conducted research in district, Bebesan, Aceh Tengah adduced that incidence of dengue related to habits of cleanliness and mosquito net usage. Rendy (2013) who conducted research in sub district, Sawah Lama, Tangerang stated that LFI (Larvae Free Index) related to all of community behavior factors, the act of closing the container and the existence of container cap only unrelated to the LFI. Aryati et al (2014) who conducted research in sub district/village Baler Bale Agung, Bali adduced that incidence of dengue related to hygienic behavior, while knowledge about the dengue unrelated to the incidence.

IPCC (2014) noted that global climate change has occurred in all aspects including the distribution of vector-bone diseases such as dengue fever. Palembang, the capital of South Sumatra province, has a tropical climate so it has a high potential for the occurrence of dengue fever. In the beginning of 2015, the city has the highest incidence of dengue fever (the Provincial Health Office of South Sumatra in Sriwijaya Post, 2015). The purpose of this paper is to map the incidence of dengue fever into five areas in Palembang based on the multinomial regression approach. Toan (2015) applied the regression to study the significant variables of dengue incidence. She showed that the incidence of dengue in Hanoi is affected by living near open sewers, detecting mosquitoes in the house, directly discharging sewage to ponds, Living in an unhygienic house, and living in rental house. These risks increased respectively by 7.9, 6.3, 4.3, 3.4, and 2.2 times. This approach also has been applied in some area research; psychology (El-Habil, 2012), image processing (Khodadadzadeh et al, 2014), and medical science (Thors´en, 2014).

2. MATERIAL AND METHODS

Data Collection and Description

The region of Palembang is separated into two by the river Musi, namely; Seberang Ulu and Seberang Ilir. We separate the spreading area of dengue fever in Palembang into 5 areas; South in the Seberang Ulu and Center, North, West, and East in the Seberang Ilir. Data collecting of the incidence of dengue fever that spread in the areas as presented in Table 1 using stratified sampling technique, where the number of samples is determined by Slovin formula. The variables considered in mapping the incidence of dengue fever is shown in Table 2.

Table 1. Proportion of incidence of dengue fever for each area.

Area	% sample
South	15.2%
Central	20.7%
North	20.0%
West	9.5%
East	34.7%

Table 2. The Independent Variables

	Subject	Predictor Variable	Notation	Type of variable
1.	A family member who get dengue fever	Age	X_1	1. Children, 2.Teens, 3. Adults
		Gender	X_2	1. Male, 2. Female
		Blood type	X_3	1. A, 2. B, 3. AB, 4. O
2.	Respondent	Age	X_4	interval
		Education level	X_5	1. not graduate primary, 2. primary, 3. secondary, 4. high school, 5. college
		Occupation	X_6	1. Housewife/Retired/ uncertain 2. Farmers/Fishermen, 3. Private Employees/SOE 4. Civil servant/army/police, 5. Self Employed
		Income per month	X_7	interval
		Knowledge of dengue fever	X_{12}	1.Good, 2. Bad
		Behavior to face the incidence of dengue fever	X_{13}	1.Good, 2. Enough, 3. Bad
3.	Head of family	Age	X_8	interval
		Education level	X_9	1. not graduate primary, 2. primary, 3. secondary, 4. high school, 5. college
		Occupation	X_{10}	1. Housewife/Retired/ uncertain 2. Farmers/Fishermen, 3. Private Employees/SOE 4. Civil servant/army/police, 5. Self Employed
		Income per month	X_{11}	interval
4.	Facilities of residence	Type of residence	X_{14}	1. Home beds, 2. Flat, 3. Staged house, 4. 5. Permanent
		Number of bedrooms	X_{15}	1. one, 2. two, 3. three, 4. four, 5. > four
		Number of bathrooms that have the tub	X_{16}	1. one, 2. two, 3. three, 4. four, 5. > four
		Source of water supply	X_{17}	1. river water, 2. rain water, 3. well water, 4. artesian well water, 5. tap water

Multinomial Logistic Regression

Consider an data of incidence of dengue fever with n respondents $(i=1,2,\cdots,n)$, five areas of the spread of dengue fever $(j=1,2,\cdots,5)$, p predictor variables $(k=1,2,\cdots,p)$, the number of sample in j-th areas is denoted n_{ij} and $n=\sum_{j=1}^{5} n_{ij}$. Let N_{ij} be the random variable for i-th respondents who have the family member experiencing dengue fever with probability into j-th area, $\pi_j(x_i)$. If N_{ij} follow multinomial logistic regression model, the probability mass function is (Resti et al, 2016)

$$f\left(n_{ij}\middle|\pi_j(x_i)\right)=\frac{n}{\prod_{j=1}^{5} n_{ij}!}\prod_{j=1}^{5}\pi_j(x_i)^{n_{ij}} \quad (1)$$

The the probability mass function as in eq (1) can be also written as

$$f(n_{i1},n_{i2},\cdots,n_{i5})=\left(1-\sum_{j=1}^{4}\pi_j(x_i)\right)^{n-\sum_{j=1}^{4}n_{ij}}\prod_{j=1}^{5}\pi_j(x_i)^{n_{ij}} \quad (2)$$

which implied that probability of n_{ij} fall into areas j-th, where $\sum_{j=1}^{5}\pi_j(x_i)=1$. The multinomial logistic regression model was formed with stated $P(N_{ij}=j|x_i)=\pi_j(x_i)$ as the following logistic functions,

$$\pi_j(x_i)=\frac{\sum_{k=0}^{p}\exp(\beta_k x_{ijk})}{1+\sum_{j=1}^{5}\sum_{k=0}^{p}\exp(\beta_k x_{ijk})} \quad (3)$$

The regression parameters, β_k , can be estimated from (1) using maximum likelihood estimation.

3. RESULT AND DISCUSSION

We explored the relationship between predictor variables and the spreading areas of dengue fever using the chi-square test and the result the variables, $X_6, X_7, X_9, X_{10}, X_{11}, X_{13}, X_{14}, X_{15}, X_{16}$ and X_{17} do not have relationship with the areas. The result test on the parameter β_k simultaneously using the likelihood ratio test/G-statistics ($G-statistic=472.95>\chi^2=391.48$) showed minimal there is one predictor variable was statistically significantly has a relationship with the dependent variable. This simultaneously test result has strengthened by the partial test results ($\alpha=5\%$) as shown in Table 3 that there are nine predictor variables was statistically significantly affect the dependent variable (all of the nine variables have $G-statistic>\chi^2$ and p-value < α).

Table 3. The partial test

Predictor Variable	G-Statistics	χ^2	p-value
Intercept	436,85	41,75	0,00
X_1	525,04	52,09	0,00
X_2	593,45	120,50	0,00
X_3	527,88	54,92	0,00
X_4	491,39	18,43	0,00
X_5	562,93	89,98	0,00
X_8	491,78	18,83	0,00
X_{12}	512,61	39,65	0,00

The decision to map the incidence of dengue fever into five areas for respondents who have a family member experiencing dengue fever is affected by family member who get dengue fever (age, gender and blood type), respondent (age, education level, and knowledge of dengue fever), and head of family (age).

The backward stepwise procedure has been applied to obtain the final model of the multinomial logistic regression as shown in Table 4, in which East area is used as the reference category and for each categorical predictor variable, the reference is the last category. The tendency of a respondent who has a male family member experiencing dengue fever would be mapped to the South rather than to the East only by 0.02 times compared to respondent who has a female family member. The tendency of the respondents have a family member experiencing dengue fever who the blood type-A and AB respectively would be mapped to the South rather than to the East by 14.91and 11.06 times compared to respondent who has the blood type-O.The younger of the respondent, the tendency to be mapped to the South of 0.91-fold than to the East. The higher the respondents' knowledge about dengue, the tendency to be mapped to the South of 5.07-fold than to the East. Based on Table 4 also, the tendency of respondents to be mapped to a area of Central, North, and West than to the East can be interpreted in the same way.

Table 4. The final model of trinomial regression

Area of dengue fever		β	p-value	exp (β)	Area of dengue fever		β	p-value	exp (β)
1. South	Intercept	0.69	0.69	-	3. North	Intercept	1.59	0.36	-
	$X_{1(1)}$	0.39	0.70	1.48		$X_{1(1)}$	-0.63	0.52	0.53
	$X_{1(2)}$	0.31	0.77	1.36		$X_{1(2)}$	-3.99	0.00	0.02
	$X_{2(1)}$	-3.74	0.00	0.02		$X_{2(1)}$	-5.50	0.00	0.00
	$X_{3(1)}$	2.70	0.00	14.91		$X_{3(1)}$	2.30	0.02	10.01
	$X_{3(2)}$	0.25	0.78	1.29		$X_{3(2)}$	-2.13	0.04	0.12
	$X_{3(3)}$	2.40	0.00	11.06		$X_{3(3)}$	0.73	0.33	2.08
	X_4	-0.09	0.12	0.91		X_4	0.10	0.17	1.10
	$X_{5(1)}$	1.39	0.21	4.01		$X_{5(1)}$	1.94	0.15	6.92
	$X_{5(2)}$	-0.45	0.58	0.64		$X_{5(2)}$	0.84	0.38	2.31
	$X_{5(3)}$	1.43	0.09	4.19		$X_{5(3)}$	3.37	0.00	28.93
	$X_{5(4)}$	-0.88	0.23	0.41		$X_{5(4)}$	2.14	0.02	8.50
	X_8	0.06	0.34	1.06		X_8	-0.12	0.11	0.89
	$X_{12(1)}$	1.62	0.01	5.07		$X_{12(1)}$	2.16	0.00	8.69
2. Central	Intercept	-1.37	0.48	-	4. West	Intercept	-0.76	0.70	-
	$X_{1(1)}$	-0.92	0.33	0.40		$X_{1(1)}$	1.39	0.17	4.03
	$X_{1(2)}$	-3.66	0.00	0.03		$X_{1(2)}$	-0.13	0.90	0.88
	$X_{2(1)}$	-3.54	0.00	0.03		$X_{2(1)}$	-1.13	0.14	0.32
	$X_{3(1)}$	1.96	0.05	7.10		$X_{3(1)}$	3.73	0.00	41.68
	$X_{3(2)}$	-0.71	0.43	0.49		$X_{3(2)}$	1.46	0.11	4.32
	$X_{3(3)}$	0.53	0.47	1.70		$X_{3(3)}$	3.53	0.00	33.99
	X_4	-0.09	0.17	0.91		X_4	0.16	0.05	1.17
	$X_{5(1)}$	1.89	0.18	6.64		$X_{5(1)}$	0.44	0.78	1.55
	$X_{5(2)}$	1.82	0.05	6.14		$X_{5(2)}$	1.47	0.09	4.35
	$X_{5(3)}$	4.85	0.00	127.69		$X_{5(3)}$	-0.09	0.93	0.92
	$X_{5(4)}$	0.58	0.56	1.79		$X_{5(4)}$	-0.91	0.29	0.40
	X_8	0.09	0.22	1.09		X_8	-0.20	0.02	0.82
	$X_{12(1)}$	3.47	0.00	31.98		$X_{12(1)}$	0.02	0.98	0.99

Based on the only significant predictor variable as shownin Table 4, the probabilities of a respondent would be mapped to area of South, Central, North, West, and East respectively at,

$$\pi_1(x_i)=\frac{\exp(-3.74x_{i1,1(2)}+2.70x_{i1,3(1)}+2.40x_{i1,3(3)}+1.62x_{i1,12(1)})}{1+\exp(-3.74x_{i1,1(2)}+2.70x_{i1,3(1)}+2.40x_{i1,3(3)}+1.62x_{i11,2(1)})+\cdots+\exp(2.70x_{i4,3(1)}+2.40x_{i4,3(3)}-0.2x_{i4,8})}$$

$$\pi_2(x_i)=\frac{\exp(-3.66x_{i2,1(2)}-3.54x_{i2,2(1)}+4.85x_{i2,5(3)}+3.47x_{i2,12(1)})}{1+\exp(-3.74x_{i1,1(2)}+2.70x_{i1,3(1)}+2.40x_{i1,3(3)}+1.62x_{i11,2(1)})+\cdots+\exp(2.70x_{i4,3(1)}+2.40x_{i4,3(3)}-0.2x_{i4,8})}$$

$$\pi_3(x_i)=\frac{\exp(-3.99x_{i3,1(2)}-5.49x_{i3,2(1)}+2.30x_{i3,3(1)}-2.13x_{i3,3(2)}+3.37x_{i3,5(3)}+2.14x_{i1,5(4)}+2.16x_{i1,12(1)})}{1+\exp(-3.74x_{i1,1(2)}+2.70x_{i1,3(1)}+2.40x_{i1,3(3)}+1.62x_{i11,2(1)})+\cdots+\exp(2.70x_{i4,3(1)}+2.40x_{i4,3(3)}-0.2x_{i4,8})}$$

$$\pi_4(x_i)=\frac{\exp(2.70x_{i4,3(1)}+2.40x_{i4,3(3)}-0.2x_{i4,8})}{1+\exp(-3.74x_{i1,1(2)}+2.70x_{i1,3(1)}+2.40x_{i1,3(3)}+1.62x_{i11,2(1)})+\cdots+\exp(2.70x_{i4,3(1)}+2.40x_{i4,3(3)}-0.2x_{i4,8})}$$

$$\pi_5(x_i)=1-\pi_1(x_i)-\pi_2(x_i)-\pi_3(x_i)-\pi_4(x_i)$$

Nagelkerke value of this final multinomial regression model is 0.784, which indicates that variance of the significant predictor variables are able to explain variance of the areas of the incidence of dengue fever at 78.4 percent, while the rest is explained by the predictor variables other than this model.

Table 5. Mapping Result

Observed	Mapped					
	South	Center	North	West	East	Percent Correct
South	17	7	11	2	6	39.5%
Center	4	32	12	3	8	54.2%
North	3	11	37	0	6	64.9%
West	4	1	0	13	9	48.1%
East	4	1	1	5	88	88.9%
Overall Percentage	11.2%	18.2%	21.4%	8.1%	41.1%	65.6%

The number of respondents who successfully mapped by the final model to area; South, Central, North, West and East respectively as many as 17, 32, 37, 13, and 88 respondent. The rest is mapped to other areas successively as many as 26, 27, 20, 14, and 11 . Overall correct percentage of respondents mapping is 65.6%. This model is good enough for mapping the incidence of dengue fever into five areas in Palembang based on both the Nagelkerke value and the correct percentage of respondents mapping.

4. CONCLUSION

Employing the multinomial regression approach in mapping the incidence of dengue fever into 5 areas; South, Central, North, West, and East, we found that the decision to map the incidence for respondents who have a family member experiencing dengue fever is affected family member who get dengue fever (age, gender and blood type), respondent (age, education level, and knowledge of dengue fever), and head of family (age). Based on the Nagelkerke value, variance of the significant predictor variables in the study able to explain variance of the areas as dependent variable at 78.4 percent, while based on the mapping result, the overall correct percentage of respondents mapping is 65.6%. This model is good enough for mapping the incidence of dengue fever into five areas in Palembang based on both statistics.

ACKNOWLEDGEMENTS

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

5. REFERENCES

Aryati, I. K. C., Sali, I. W., Aryasih, I. G. A. M., 2014, Hubungan Pengetahuan, Sikap, dan Tindakan Masyarakat terhadap Kejadian Demam Berdarah. *Jurnal Kesehatan Lingkungan*, 4(2): 118-123.

El-Habil, A., 2012, An Application on Multinomial Logistic Regression Model, Pakistan Journal of Statistics and Operation Research. Vol.VIII No.2, pp.271-291.

Intergovernmental Panel on Climate Change, 2014. Summary for Policymakers: Climate change 2014. Impacts, Adaptation, and Vulnerability. Technical report.

Knowlton, K., Solomon, G., and Ellman, M.R, 2009, Mosquito-Borne Dengue Fever Threat Spreading in the Americas, *The Natural Resources Defense Council Paper*, July.

Khodadadzadeh, M., Li, J., Plaza, A., and Bioucas-Dias, J. M., 2014, A Subspace-Based Multinomial Logistic Regression for Hyperspectral Image Classification, IEEE Geoscience and Remote Sensing Letters, Vol. 11, No. 12, December, pp. 2105-2109.

Media Indonesia, 2016, http://www.mediaindonesia.com/news/read/76527/kasus-dbd-di-sumsel-meningkat-23-orang-meninggal/2016-11-09.

Rendy, M. P., 2013, Hubungan Faktor Perilaku dan Faktor Lingkungan dengan Keberadaan Larva Nyamuk Aedes Aegypti di Kelurahan Sawah Lama Tahun 2013, *Skripsi*, Peminatan Kesehatan Lingkungan. Prodi Kesehatan Masyarakat. FKIK. UIN Syarif Hidayatullah.

Resti, Y., Mohruni, A.S., Burlian, F., Yani., and Amran, Ali, 2016, A probability approach in cans identification, Sriwijaya International Conference on Engineering, Science, and Technology, 8-10 November, 2016.

Sari, D., 2012, Hubungan Pengetahuan dan Perilaku Responden terhadap Kejadian Demam Berdarah di kecamatan Bebesen, Kabupaten Aceh Tengah 2012. *Skripsi*, Universitas Indonesia.

Sriwijaya Post, 2015, http://palembang.tribunnews.com/2015/02/02/palembang-tertinggi-penderita-demam-berdarah

Thors´en, E., 2014, Multinomial and Dirichlet-multinomial Modeling of Cate.

Toan, D.T.T., 2015, Climate Change and Dengue Transmission in Vietnam: an Integrated Assessment, Dissertation of Maastricht University, Netherlands.

The Mapping of Incidence of Dengue Fever in Palembang: A Multinomial Regression Aprroach

ORIGINALITY REPORT

5%
SIMILARITY INDEX

PRIMARY SOURCES

1	mafiadoc.com Internet	64 words — 2%
2	www.neliti.com Internet	35 words — 1%
3	www.nrdc.org Internet	22 words — 1%
4	link.springer.com Internet	20 words — 1%

EXCLUDE QUOTES	OFF	EXCLUDE SOURCES	< 1%
EXCLUDE BIBLIOGRAPHY	ON	EXCLUDE MATCHES	< 1 WORDS