International Journal of Applied Business and Information Systems ISSN: 2597-8993
Vol 2, No. 1, March 2018 pp. 11-17

Application of Spatial Regression Model for Modeling Measles Case in Indonesia

[1] Tuti Purwaningsih, [2] Mutiara Herawati, [3] Nanda HadinaWijayanti

[1,3]Department of Statistics, Universitas Islam Indonesia, Yogyakarta
[2]Department of Pharmacy, Universitas Islam Indonesia, Yogyakarta

[1] tuti.purwaningsih@uii.ac.id [2]mutiara.herawati@uii.ac.id [3]14611106@students.uii.ac.id

ARTICLE INFO

Article history:
Received December 20, 2017
Revised January 20, 2018
Accepted February 15, 2018

Keywords:
Measles, SEM, SLM, Factors

ABSTRACT

Measles is also known as morbili in Latin and measles in English. Measles, in the past is considered as something that must be experienced by every child, they assume, that measles can heal itself if it was already out, so that children with measles do not need to be treated. This study examines the case of measles and the causes of measles. The variables used in the study were cases of measles (Y), population density (X1), immunization coverage (X2), average incidence (X3), and number of deaths (X4) in Indonesia covering all provinces. The study examined the pattern of spread, then given a SEM application to identify how much influence the measles factor can affect the case of measles in Indonesia. The results of the study show that Measles Cases in Indonesia have a regional grouping pattern. The modeling results using SEM show lambda and all significant variables. The SEM model produced AIC of 462,429 which was better than the regression of the SLM model with AIC of 467,499.

I. Introduction

Health is a matter that must be considered. Not only for yourself, health also plays an important role in the progress of a nation. However, there are several health problems faced by various countries in the world. One of them is measles. This disease is very potential to cause an outbreak. In medical language, measles is known as morbili. Measles is a contagious respiratory disease. The measles virus is transmitted through the air when the person coughs or sneezes. Measles is caused by a viral infection that lives in the fluid of the lenders in the nose, throat and in the blood.

Immunization-preventable diseases still threaten the world, including in Indonesia. This is evident that until 2014, measles outbreaks still occurred in Indonesia. Although the trend is declining, the numbers are still high, Measles Outbreaks occurred in 10,651 cases compared to 2013 with 18,488 cases.

The measles immunization program in Indonesia began in 1982. Indonesian children aged 1-2 years who received measles immunization reached an average of 74.4%. Whereas, the achievement of measles immunization in Indonesia until December 2013 was 90.82% [4]. Although measles immunization outcomes in Indonesia have covered 90%, WHO reported that there were around 6,300 cases of measles in Indonesia in 2013. measles immunization coverage was high (> 90%) but there were still children affected by measles [10]. Because the remaining 10 percent of children who have not received immunization coupled with 10 percent of children immunized but not immune causes the community's immunity to only reach 81%.

Community perceptions of measles immunization can be different, depending on the way a person understands and understands the importance of immunization measures performed on his child. Based on research, 75% of mothers agreed that their children were immunized because they did not want to deviate from existing norms and cultures and wanted to comply with existing rules. The rest, 25% of mothers, delay giving immunization to their babies or refusing their children to be immunized because they have a perception that they are not sure that the vaccine can prevent the disease and feel worried about the side effects caused by the vaccine [1].

Based on research on measles in Indonesia in general which causes cases of measles in Indonesia to increase, the socio-economic conditions, immunization coverage and population density. So in this study taking 3 factors that influence the case of measles in Indonesia, namely the number of population, population density, number of poor people, average expenditure per capita a month and coverage of basic immunization.

II. Literature Review

A. Measles

Measles is a very contagious acute disease caused by a virus. Measles is also called rubeola, morbili, or measles. This disease is transmitted through droplets or contact with patients. This disease has an incubation period of 8-13 days. Measles is characterized by early symptoms of fever, cough, runny nose, and conjunctivitis which are then followed by reddish spots on the skin (rash). The impact of measles in the future is malnutrition as a result of recurrent and prolonged diarrhea after measles, brain inflammation syndrome in children over 10 years, and pulmonary tuberculosis becomes more severe after severe measles.

B. Spatial Regression Model

Spatial Regression is a method in modeling a data that has spatial elements. The general model of spatial regression or usually also called Autoregressive Moving Average (SARMA) in the form of a matrix [2] [6] [9] [12]. can be presented as follows:

$$Y = \rho Wy + X\beta + u$$

$$u = \lambda Wu + \varepsilon$$

$$\varepsilon \sim N(0, \sigma^2 I)$$

C. Spatial Error Model (SEM)

SEM is a spatial model that occurs due to the spatial effect on errors. If the data obtained results in lag dependencies, the data is modeled with SAR, but if the data produces error dependencies, the data is modeled with SEM. If the data produces lag dependencies and error dependencies, the data is modeled with Spatial Autoregressive Moving Average (SARMA). To see the closeness of the relationship between one region and another region in spatial data can be used spatial weighting matrix. The types of spatial weighting matrices include edge contiguity, rook contiguity, bishop contiguity, and queen contiguity.
The general form of the SEM equation [6] [7] [12] is as follows:

$$\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+(\mathbf{I}-\lambda\mathbf{W})-1\boldsymbol{\varepsilon} \text{ (2.8) } \boldsymbol{\varepsilon}\sim N(0,\mathbf{I}\sigma 2)$$

The form of estimating parameters from the SEM regression model, namely as follows:

$$\boldsymbol{\beta}=[(\mathbf{X}-\lambda\mathbf{WX})t(\mathbf{X}-\lambda\mathbf{WX})]-1(\mathbf{X}-\lambda\mathbf{WX})t(\mathbf{I}-\lambda\mathbf{W})\mathbf{y}$$

D. Selection of the Best Models

The selection criteria for the model used in this study are:

a. Coefficient of Determination (R2)
Notated with R2 = (18)
with :
SSR : Sum Square Regression (Number of squares of regression)
SST : Sum Square Total (Total Number of Squares)
The greater the value of R2 indicates the greater confidence in the model.

b. Akaike Info Criterion (AIC)
Notated with AIC = -2Lm + 2m (19)
where Lm= Maximum log-likelihood
m = number of parameters in the model.

Models with small values are the best [8]

III. Research Method

A. Data Source

In this study the data used are secondary data obtained from Indonesia Bureau of Statistics (BPS) and the Indonesian Health Profile in 2014 [5] [11]. These data include cases of measles, population density, immunization coverage, average incidence, and the number of deaths in Indonesia which includes all provinces.

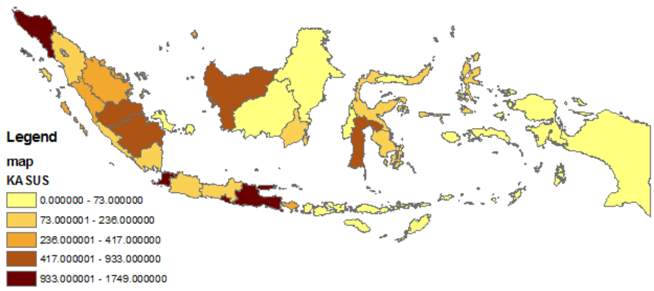
B. Research Variable

In this study there is one dependent variable and 4 independent variables, namely:

a. Number of measles cases in Indonesia (Y)
b. Population density (X1)
c. Immunization coverage (X2)
d. Average event (X3)
e. The number of victims died (X4)

IV. Results and Discussion

The following is a demographic pattern of spread of measles cases and factors affecting measles:

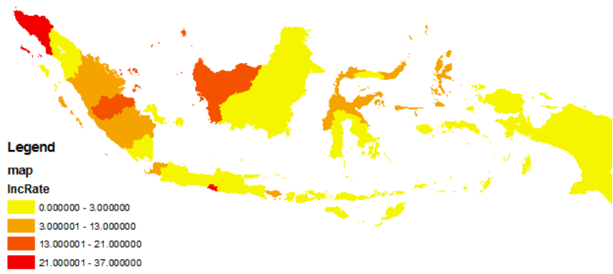


Source: Data processed

Fig. 1: Distribution of Measles Cases in Indonesia by Province in 2014

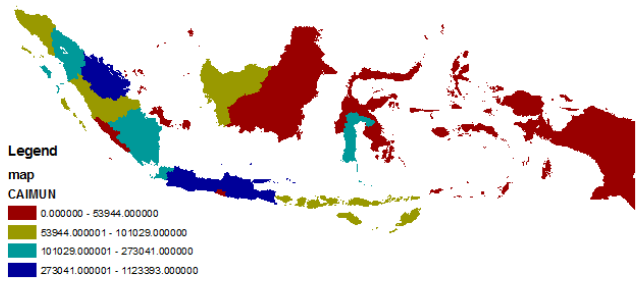
Source: Data processed

Fig. 2: Distribution of Victims Who Died in Indonesia by Province in 2014



Source: Data processed

Fig. 3: Average Distribution of Measles in Indonesia by Province in 2014



Source: Data processed

Fig. 4: Distribution of Immunization Coverage in Indonesia by Province in 2014

Source: Data processed

Fig. 5: Distribution of Population Density in Indonesia by Province in 2014

Regression Model

Spatial Error Model (SEM)

Result of SEM Model:

```
SUMMARY OF OUTPUT: SPATIAL ERROR MODEL - MAXIMUM LIKELIHOOD ESTIMATION
Data set              : map
Spatial Weight        : map.gal
Dependent Variable    :        KASUS  Number of Observations:   33
Mean dependent var    :  387.515152  Number of Variables   :    5
S.D. dependent var    :  460.179288  Degrees of Freedom    :   28
Lag coeff. (Lambda)   :   -0.431096

R-squared             :    0.767003  R-squared (BUSE)      : -
Sq. Correlation       : -            Log likelihood        : -226.214459
Sigma-square          :     49340.6  Akaike info criterion :    462.429
S.E of regression     :     222.127  Schwarz criterion     :    469.911
```

Variable	Coefficient	Std.Error	z-value	Probability
CONSTANT	23.76107	48.6003	0.4889078	0.62491
KEPPEND	0.05493968	0.01519409	3.615858	0.00030
CAIMUN	0.0004899284	0.0001581835	3.097215	0.00195
IncRate	37.68267	4.611961	8.170638	0.00000
MENINGGAL	-118.9394	67.06327	-1.773541	0.07614
LAMBDA	-0.4310962	0.1812415	-2.378573	0.01738

The Next is modeling using SEM [3]. Based on the Geoda output from above SEM result, it shows that from keppend (X1), caimun (X2), the inncrate (X3) has a positive sign while the residence is negative and significant at the 5% level. The effect of spatial correlation is accommodated in the model by entering the lambda spatial weighing variable so that lambda plays an important role in SEM modeling. In addition, the Keppend, caimun, incrate and death variables play an important role in SEM with a significance level of 5%. This means that canpak cases in an area are influenced by the value of Keppend, caimun, incrate and death by 76.7% while the remaining 23.3% is explained by other variables outside the model.

The SEM model formed is as follows.

$$\hat{Y} = 23.76107 + 0.0549\mathrm{X1} + 0.0004899\mathrm{X2} + 37.68267\mathrm{X3}$$
$$- 188.9394\mathrm{X4} - 0.4310 \sum_{j=1, i\neq j}^{n} w_{ij} y_j$$

The SEM model can be interpreted that the effect of kepend on cases is the same for each province at 0.0549. This means that if other factors are considered constant, if the keppend value in a province rises by 1 unit then the value of the case will increase by 0.0549 units. The effect of caimun on cases is also the same for each province at 0,0004899. This means that if other factors are considered constant, if the value of caimun in a province rises by 1 unit then the HDI value will increase by 0,0004899 units. The effect of incrate on cases is the same for each province with its elasticity of 37.68267. This means that if other factors are considered constant, if the incrate value in a province rises by 1 unit then the value of the case will increase by 37,68267 units. Similarly, the effect of death on cases is the same for each province with its elasticity of -188.9394. This means that if other factors are considered constant, if the incrate value in a province increases by 1 unit then the value of the case will increase by -188.9394 units.

Spatial Lag Model (SAR)

Result of the SAR model:

```
SUMMARY OF OUTPUT: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION
Data set              : map
Spatial Weight        : map.gal
Dependent Variable    :       KASUS  Number of Observations:   33
Mean dependent var    :     387.515  Number of Variables   :    6
S.D. dependent var    :     460.179  Degrees of Freedom    :   27
Lag coeff.   (Rho)    :  -0.0273959

R-squared             :    0.727027  Log likelihood        :    -227.749
Sq. Correlation       : -            Akaike info criterion :     467.499
Sigma-square          :     57806.1  Schwarz criterion     :     476.478
S.E of regression     :     240.429

-----------------------------------------------------------------------
    Variable    Coefficient     Std.Error       z-value    Probability
-----------------------------------------------------------------------
     W_KASUS   -0.02739595      0.1448624    -0.1891171      0.85000
    CONSTANT      10.61367        67.4662     0.1573183      0.87499
     KEPPEND    0.05100085     0.01675897      3.043197      0.00234
      CAIMUN 0.0005018631   0.0002224757      2.255811      0.02408
     IncRate      39.94197       4.930185      8.101515      0.00000
   MENINGGAL     -121.1067       80.47996     -1.504806      0.13237
-----------------------------------------------------------------------
```

Next is modeling using the Spatial Lag Model (SLM) [3]. The following are the outputs of the analysis process:

$$\hat{Y} = -0.027 \sum_{j=1,i\neq j}^{n} w_{ij}y_j + 0.61367 + 0.0510X1 + 0.00050X2 + 39.94197X3 - 121.1067X4$$

The above model can be interpreted: The number of measles cases in the 4th province will increase by 0.027 if there is an increase in cases of measles by 1 unit in the j province with other factors being considered constant. The number of measles cases in the second province will increase by 0.0510 if the population density in the i province increases by 1 unit with other variable requirements still. The number of measles cases in the 4th province will increase by 0,00050 if immunization coverage in the first province has increased by 1 with other variable conditions remaining. The number of cases of measles in the province of the i will decrease by 121,106 if there is a death toll in the province of i experiencing an increase of 1 unit with other variable conditions remaining. The number of measles cases in the 4th province will increase by 39.94197 if the average incidence of measles in the first province has increased by 1 with other variable conditions remaining.

Comparison of SLM Models and SEM Models

The goodness criteria of the model used is to compare the AIC values of the two models.

Model	AIC
SEM	462.429
SLM	467.499

Based on the table above, it can be seen that the model with the minimum AIC value is the SEM model. So that the SEM model is better used to analyze data on the number of measles cases in Indonesia compared to the SLM model.

Based on the relationship between the number of measles cases in Indonesia with population density, immunization coverage, average incidence, and the number of deaths, it can be interpreted that the similarities and differences in characteristics of each Provinces cause an increase or decrease in the number of measles cases in Indonesia.

V. Conclusion

The pattern of the spread of measles cases in Indonesia seems to be clustered between adjacent areas. Based on the relationship between cases of measles in Indonesia with population density, immunization coverage, average incidence, and the number of deaths, it can be interpreted that similarities and differences in characteristics in each adjacent province can lead to an increase or decrease in measles cases in Indonesia. SEM Regression Model is better than SLM. The SEM model formed to model measles cases in Indonesia in 2014 was:

$$\hat{Y} = 23.76107 + 0.0549\text{X1} + 0.0004899\text{X2} + 37.68267\text{X3} - 188.9394\text{X4} - 0.4310 \sum_{j=1, i \neq j}^{n} w_{ij} y_j$$

VI. Reference

[1] Andrea et al. (2006). Qualitative Analysis of Mother's Decision-Making about Vaccines for Infants: The Importance of Trust. Pediatrics, Vol. 117/Issue 5

[2] Anselin, L. (2004). Advances in Spatial Science. Advances in Spatial Econometrics: Methodology, Tools, and Applications, 383–395. https://doi.org/10.1007/978-3-642-17940-2

[3] Anselin, L., Syabri, I., & Kho, Y. (2006, January). GeoDa: An introduction to spatial data analysis. Geographical Analysis. https://doi.org/10.1111/j.0016-7363.2005.00671.x

[4] Badan Penelitian dan Pengembangan Kesehatan KEMENKES. (2010). Riset Kesehatan Dasar. 2010.Sumber: http://kesga.kemkes.go.id/images/pedoman/Riskesdas%202010%20Nasional.pdf

[5] Badan Pusat Statistik. (2014). http://www.bps..go.id

[6] Beale, C. M., Lennon, J. J., Yearsley, J. M., Brewer, M. J., & Elston, D. A. (2010, February). Regression analysis of spatial data. Ecology Letters. https://doi.org/10.1111/j.1461-0248.2009.01422.x

[7] Elhorst, J. P. (2010). Applied Spatial Econometrics: Raising the Bar. Spatial Economic Analysis, 5(1), 9–28. https://doi.org/10.1080/17421770903541772

[8] Hoeting, J. A., Davis, R. A., Merton, A. A., & Thompson, S. E. (2006). Model selection for geostatistical models. In Ecological Applications(Vol. 16, pp. 87–98). https://doi.org/10.1890/04-0576

[9] LeSage, J. P. (1999). Applied Econometrics using MATLAB. Review Literature And Arts Of The Americas (Vol. 204, pp. 1–332). https://doi.org/10.1371/journal.pone.0052763

[10] Pusat komunikasi Publik KEMENKES RI. (2015). KLB Difteria dan Campak Masih Terjadi!. Sumber: http://mediakom.sehatnegeriku.com/klb-difteria-dan-campak-masih-terjadi/

[11] Kementrian Kesehatan RI. (2014). Profil Kesehatan Indonesia. http://www.depkes.go.id/resources/download/pusdatin/profil-kesehatan-indonesia/profil-kesehatan-indonesia-2014.pdf

[12] Schabenberger, O., & Gotway, C. A. (2017). Statistical methods for spatial data analysis. Statistical Methods for Spatial Data Analysis (pp. 1–488). CRC Press. https://doi.org/10.1201/9781315275086