Vol 01 No 2 2022, pp 20-27
e-ISSN: XXXX-XXXX
DOI: https://doi.org/10.54482/IJEBIIITS/vol01-iss1/5

Spatial Regression Analysis of 2020 Pulmonary TB Cases in North Sumatra Province

Aulia Fajrin Ramadhani Nasution[a*], Defi Afriyanti[a], Ainiyah Imaarah[a], Hafiz Jannesman[a], Mutiara Zenita[a]
[a]Department of Geography, Faculty of Social Science, State University of Padang
*Coresponding email: arinnasution55@gmail.com

ABSTRACT

Tuberculosis is a disease caused by a contagious infection of the bacterium Mycobacterium tuberculosis which is transmitted through droplets in the air. Based on information from the Indonesian Ministry of Health, there has been an increase in cases since 2017 by 254 cases per 100,000 population, in 2018 250 cases per 100,000 population and in 2019 to 254 cases per 100,000 population in Indonesia. Based on a report from the Central Statistics Agency of North Sumatra, in 2018 cases of pulmonary TB in North Sumatra Province as a whole reached 23,237 cases, while in 2020 there were around 33,779 cases of the spread of this disease. This indicates an increase in cases of 10,542 cases in the span of two years. This study aims to provide an overview related to the distribution pattern of the number of pulmonary TB patients in the North Sumatra Province area by utilizing technology based on geographic information systems using spatial statistical regression method. The applications used in this processing include ArcGIS in processing data and GeoDa in regression analysis. The results showed that there was a correlation between the population density variable between regions and the level of pulmonary TB sufferers in North Sumatra. Based on the test statistic value in the Spatial Autoregressive Model is 2.5634 with a p-value of 0.473466 indicating the decision to reject H0 which states that there is a spatial dependence between regions with population density variables and the high and low cases of pulmonary TB that occur in districts and cities in North Sumatra Province and supported with the results of the autocorrelation Moran's I index obtained is 0.268999 with a p-value of 0.000001 so that it shows that there is a positive spatial autocorrelation.

Keywords: Pulmonary TB, North Sumatra Province, Spatial Regression

1. INTRODUCTION

Tuberculosis disease or also known as TB is a disease caused by an infectious infection of the rod-shaped Mycobacterium tuberculosis bacteria and is transmitted through droplets in the air (Irwan, 2018). The World Health Organization (WHO) in 2020 stated that in the last three years Indonesia was ranked second as the country with the most TB cases in the world (WHO, 2019). At least there has been an increase in cases since 2017 by 254 cases per 100,000 population, in 2018 250 cases per 100,000 population, and in 2019 to 254 cases per 100,000 population (Ministry of Health 2019, in Afifani, 2021). Currently, pulmonary TB is still a major health problem in Indonesia and globally it is also an important health issue to pay attention to because the spread of this type of disease tends to be easy and difficult to detect early.

North Sumatra Province is one of the largest contributors to the number of pulmonary TB patients in Indonesia. Based on a report from the Central Statistics Agency of North Sumatra, in 2018 cases of pulmonary TB in North Sumatra Province as a whole reached 23,237 cases, while in 2020 there were around 33,779 cases of the spread of this disease. This indicates an increase in cases of 10,542 cases in the span of two years (BPS, 2022). John Gordon said that infectious diseases, such as pulmonary TB, are influenced by three main factors, including the disease as an agent, the host as the host, and the environment (Ramadhani, 2021). Factors from the risk of spreading pulmonary TB disease are also related to the environment, in this discussion they are summarized in regional bandages (Hartanto, 2018).

One of the delivery of information based on the spatial conditions of the region can utilize spatial-based technology, namely the Geographic Information System (GIS) combined with spatial statistical analysis. The data analyzed based on statistical patterns with spatial analysis were able to provide an overview related to the potential distribution of the number of pulmonary TB patients in a regional unit, both based on the aspect of regional proximity and the distribution pattern based on the variable number of patients compared to population density per region. This study aims to provide an overview related to the distribution pattern of the number of pulmonary TB patients in the North Sumatra Province by utilizing geographic information system-based technology using spatial statistical regression method. The applications used in this processing include ArcGIS in processing data and GeoDa in regression analysis

2. METODOLOGI PENELITIAN

The method used in this research is the method of spatial statistics with regression analysis. Spatial regression is an analysis that evaluates the relationship between several variables by giving a spatial effect on several locations that become the center of observation. Spatial regression in this study uses an area approach with the level of pulmonary TB disease recorded by the Central Statistics Agency of North Sumatra Province in 2020.

In the spatial regression model, there are 2 spatial effects, namely spatial dependence and spatial heterogeneity (Anselin 1988). According to Anselin (1988), the general form of the spatial regression model is as follows:

$$\mathbf{y = \rho W y + X\beta + u}$$

$$\mathbf{u = \lambda W u + \varepsilon}$$

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

y = response variable vector of size n×1

X = explanatory variable matrix of size n × (p + 1)

β = is a regression parameter coefficient vector measuring (p+1)×1

W = spatial weighting matrix of size n × n

u = error vector which is assumed to contain autocorrelation of size n × 1

ε = autocorrelation-free error vector of size n × 1

λ = spatial lag autoregression coefficient, is the spatial error autoregression coefficient

I = identity matrix of size n × n

The spatial regression model is divided into two, including the Spatial Autoregressive Model or also known as the Spatial Autoregressive Model (SAR), where if the value of 0 and = 0 then the model is a SAR (Spatial Autoregressive Regression) model. This means that the model has a spatially correlated response variable. The spatial regression model becomes (Bivand et al. 2008):

$$\mathbf{y = \rho W y + X\beta + \varepsilon}$$

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

$$\mathbf{\varepsilon = y - \rho W y - X\beta}$$

$$\mathbf{= (I - \rho W)y - X\beta}$$

y = response variable, is the explanatory variable matrix

X = spatial weighting matrix, and is the predictor coefficient of the spatial lag model.

Another regression model is the Spatial Error Model or also known as the Spatial Error Model (SEM). This model is used if the value of 0 and = 0 then the model is a SEM (Spatial Error Model) model. This means that the model has a spatially correlated error variable. The spatial regression model becomes:

$$\mathbf{y = X\beta + u}$$

$$\mathbf{u = \lambda W u + \varepsilon}$$

$$\mathbf{\varepsilon = (I - \lambda W)(y - X\beta)}$$

3. RESULTS AND DISCUSSION

The number of pulmonary TB cases in each district / city in North Sumatra Province in 2020 according to the Central Bureau of Statistics of North Sumatra, which is presented in the following tables and figure:

Table 1. The number of pulmonary TB case

Kabupaten / Kota	Kasus TBC	Jumlah Penduduk	Percentage
Asahan	532	735026	0.072378392
Batu Bara	462	420103	0.10997303
Dairi	548	285481	0.191956733
Deli Serdang	3326	2234320	0.148859608
Gunungsitoli	222	143776	0.154406855
Humbang Hasundutan	283	191776	0.147567996
Karo	828	421997	0.196209926
Kota Binjai	3326	279302	1.190825701

Kota Medan	12105	2295003	0.527450291
Kota Tanjungbalai	449	177005	0.253665151
Labuhanbatu	1533	501596	0.305624447
Labuhanbatu Selatan	460	344819	0.133403322
Labuhanbatu Utara	488	366603	0.133114023
Langkat	1450	1048100	0.138345578
Mandailing Natal	885	451028	0.196218417
Nias	173	143983	0.120153074
Nias Barat	590	82425	0.715802244
Nias Selatan	167	322520	0.051779735
Nias Utara	16	138800	0.011527378
Padang Lawas	774	286627	0.270037366
Padang Lawas Utara	452	277423	0.162928092
Padangsidimpuan	564	224483	0.251243969
Pakpak Barat	140	49688	0.281758171
Pematangsiantar	784	257110	0.304927852
Samosir	244	126710	0.192565701
Serdang Bedagai	929	617772	0.150379104
Sibolga	463	87791	0.527388912
Simalungun	1718	871678	0.197091128
Tapanuli Selatan	660	283389	0.232895419
Tapanuli Tengah	806	382917	0.210489479
Tapanuli Utara	581	303688	0.19131477
Tebingtinggi	235	166100	0.141481036
Toba Samosir	456	184493	0.247163849

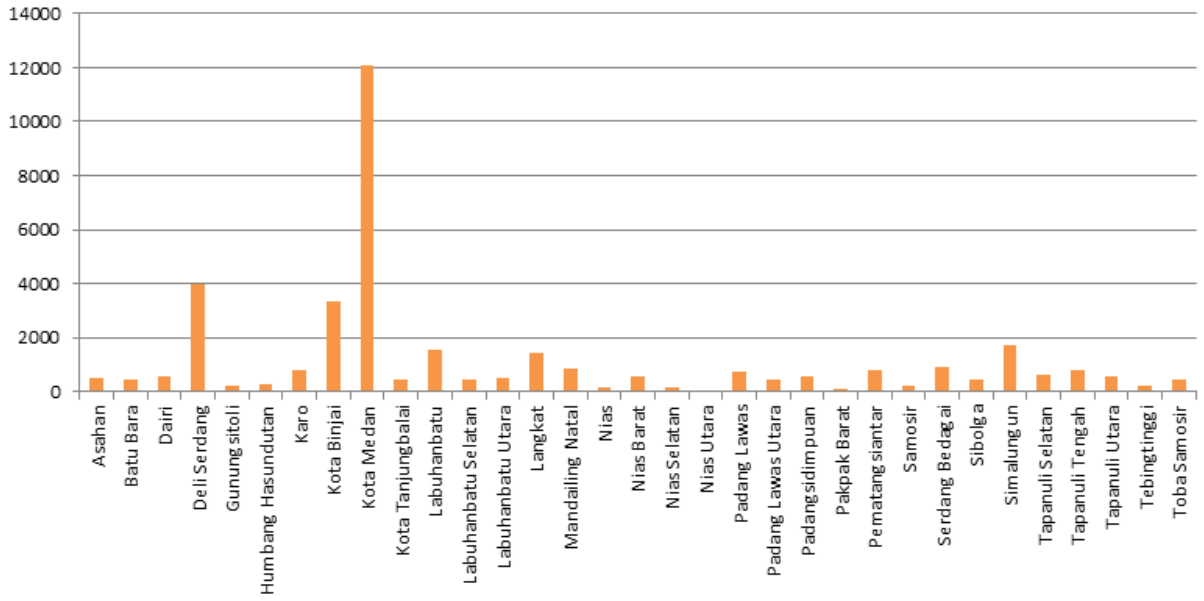


Figure 1. The number of pulmonary TB Case

The spread of this disease covers 34 regencies/cities in North Sumatra where the highest cases are in Medan City with 12105 cases, followed by Deli Serdang Regency with 3960 cases, and Binjai City with 3326 cases.

The number of TB cases that occurred in North Sumatra Province in 2020 had an influence on its surrounding neighbors. This is known through the value of the Moran's I index which has been tested previously. The value of the Moran's I index obtained is 0.268999 with a p-value of 0.000001 so that it shows that there is a positive spatial autocorrelation

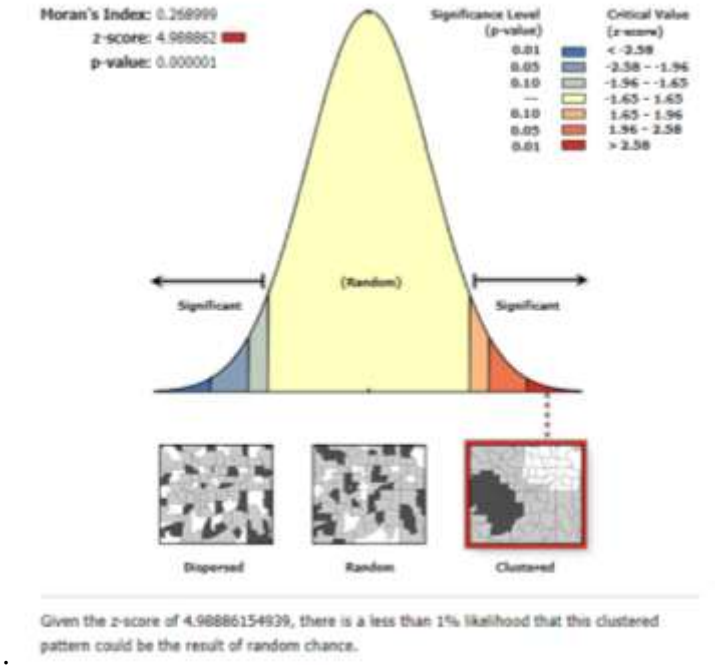


Figure 2. The number of TB case in North Sumatra in 2020

The map of results from the Moran test shows that there is a concentration on the spread of pulmonary TB in North Sumatra Province, this indicates that there is a true regional spatial influence on cases of pulmonary TB spread, especially in North Sumatra Province.

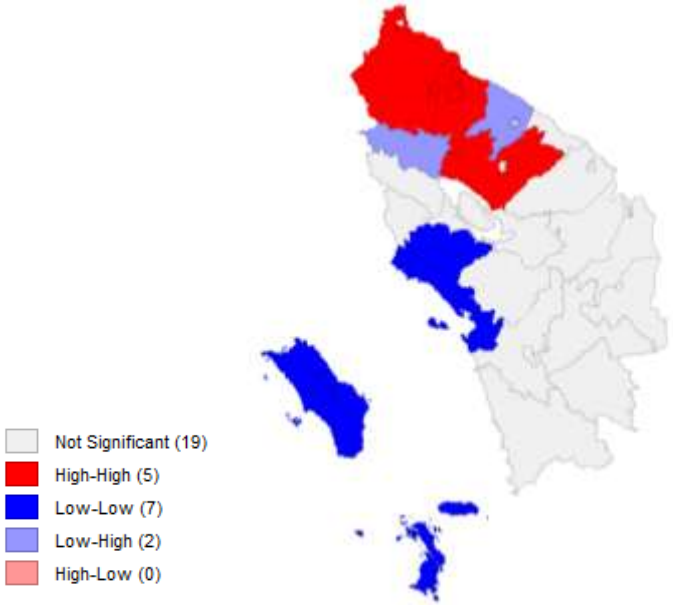


Figure 3. The map of Results from the Moran test

Regression Test Results:

a) SAR Model

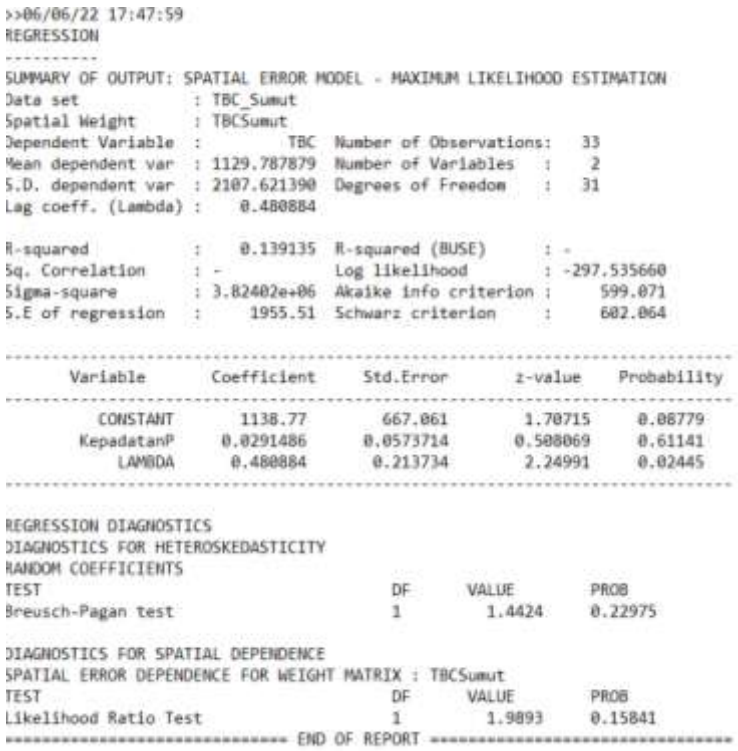
Based on the test statistic value in the SAR model is 2.5634 with a p-value of 0.473466 indicating the decision to reject H0 which states that there is a spatial dependence in the lag and it is necessary to continue with the formation of a SAR model.

```
>>06/06/22 17:47:51
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION
Data set            : TBC_Sumut
Spatial Weight      : TBCSumut
Dependent Variable  :        TBC  Number of Observations:   33
Mean dependent var  :    1129.79  Number of Variables   :    3
S.D. dependent var  :    2107.62  Degrees of Freedom    :   30
Lag coeff.   (Rho)  :   0.473466

R-squared           :   0.152826  Log likelihood        :  -297.249
Sq. Correlation     : -           Akaike info criterion :   600.497
Sigma-square        :  3.7632e+06 Schwarz criterion     :   604.987
S.E of regression   :      1939.9

------------------------------------------------------------------------------
      Variable       Coefficient     Std.Error       z-value     Probability
------------------------------------------------------------------------------
           W_TBC      0.473466       0.214512        2.20717     0.02730
        CONSTANT       524.272        436.123        1.20212     0.22932
      KepadatanP     0.0511341      0.0553933        0.92311     0.35595
------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                                DF      VALUE         PROB
Breusch-Pagan test                  1         3.0684      0.07983

DIAGNOSTICS FOR SPATIAL DEPENDENCE
SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : TBCSumut
TEST                                DF      VALUE         PROB
Likelihood Ratio Test               1         2.5634      0.10936
================================ END OF REPORT =================================
```

Figure 4. SAR Model

b) SEM Model

The value of the test statistic on the SEM model is 1.9893 with a p-value of 1138.77 so that this result gives a decision to accept H0 which states that there is no spatial dependence in the error.



```
>>06/06/22 17:47:59
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIAL ERROR MODEL - MAXIMUM LIKELIHOOD ESTIMATION
Data set            : TBC_Sumut
Spatial Weight      : TBCSumut
Dependent Variable  :         TBC  Number of Observations:   33
Mean dependent var  : 1129.787879  Number of Variables   :    2
S.D. dependent var  : 2107.621390  Degrees of Freedom    :   31
Lag coeff. (Lambda) :    0.480884

R-squared           :    0.139135  R-squared (BUSE)      : -
Sq. Correlation     : -            Log likelihood        : -297.535660
Sigma-square        : 3.82402e+06  Akaike info criterion :    599.071
S.E of regression   :     1955.51  Schwarz criterion     :    602.064

------------------------------------------------------------------------------
      Variable       Coefficient     Std.Error       z-value    Probability
------------------------------------------------------------------------------
         CONSTANT       1138.77        667.061        1.70715     0.08779
       KepadatanP     0.0291486      0.0573714       0.508069     0.61141
           LAMBDA      0.480884       0.213734        2.24991     0.02445
------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                                DF      VALUE        PROB
Breusch-Pagan test                  1         1.4424     0.22975

DIAGNOSTICS FOR SPATIAL DEPENDENCE
SPATIAL ERROR DEPENDENCE FOR WEIGHT MATRIX : TBCSumut
TEST                                DF      VALUE        PROB
Likelihood Ratio Test               1         1.9893     0.15841
================================ END OF REPORT =================================
```

Figure 5. SAM Model

```
------------------------------------------------------------------------------
      Variable       Coefficient     Std.Error       z-value    Probability
------------------------------------------------------------------------------
          W_TBC       0.473466       0.214512        2.20717     0.02730
       CONSTANT        524.272        436.123        1.20212     0.22932
     KepadatanP     0.0511341      0.0553933       0.92311     0.35595
------------------------------------------------------------------------------
```

Figure 6. SAR Model

The population density variable has a coefficient with a positive sign so that it illustrates that if there is an increase in TB cases in a district/city, it can be caused by the population density factor. The increasing population density will result in intense interactions between individuals and other community individuals, thereby increasing the possibility of increasing the spread of the TB virus through the air. A significant p-coefficient indicates that a district/city area that has high TB cases is surrounded by districts/municipalities that also have high TB cases. Vice versa.

4. CONCLUSION

Based on a report from the Central Statistics Agency of North Sumatra, in 2018 cases of pulmonary TB in North Sumatra Province as a whole reached 23,237 cases, while in 2020 there were around 33,779 cases of the spread of this disease. This indicates an increase in cases of 10,542 cases in the span of two years. The spread of this disease covers 34 regencies/cities in North Sumatra where the highest cases are in Medan City with 12105 cases, followed by Deli Serdang Regency with 3960 cases, and Binjai City with 3326 cases. The number of TB cases that occurred in North Sumatra

Province in 2020 had an influence on surrounding neighbors and was validated by the Moran's I index value obtained was 0.268999 with a p-value of 0.000001 so that it showed that there was a positive spatial autocorrelation. The right method for this regression analysis is the Spatial Autoregressive Model. Based on the test statistic value in the SAR model is 2.5634 with a p-value of 0.473466 indicating the decision to reject H0 which states that there is a spatial dependence between spatial regions.

REFERENCES

Afifiani, K., Wardani, R. S., & Kristini, T. D. (2021, December). Pola Spasial Sebaran Kasus Baru Tuberkulosis Paru. In Prosiding Seminar Nasional Unimus (Vol. 4).

Irwan.2018. Epidemiologi Penyakit Menular. Yogyakarta : CV Absolute Media.

Kemenkes RI. 2019. Profil Kesehatan Indonesia Tahun 2019. Jakarta: Kemenkes

Hikma, F., Amareta, D. I., & Maharani, H. E. (2016). Pemetaan Persebaran Penyakit Tuberkulosis Di Kabupaten Jembertahun 2013-2015. Jurnal Manajemen Informasi Kesehatan Indonesia (JMIKI), 4(1).

Hartanto, T. D., Saraswati, L. D., Adi, M. S., & Udiyono, A. (2019). Analisis Spasial Persebaran Kasus Tuberkulosis Paru Di Kota Semarang Tahun 2018. Jurnal Kesehatan Masyarakat (Undip), 7(4), 719-727.

Ramadhani, M. S., Suhartono, S., & Setiani, O. (2021). GAMBARAN SEBARAN KASUS TUBERKULOSIS DALAM PENDEKATAN SPASIAL DAN TEMPORAL. Jurnal Kesehatan Masyarakat (Undip), 9(4), 529-540.