

Pemodelan Jumlah Kasus Tuberkulosis di Indonesia Menggunakan Regresi Spasial Berbasis Area

Mar Atun Solihah*

Abstrak

Komitmen Indonesia dalam mencapai eliminasi tuberkulosis (TB) tahun 2030 yaitu menurunkan insiden TB menjadi 65/100.000 penduduk. Tahun 2021 jumlah kasus TB ditemukan sebanyak 146/100.000 penduduk meningkat dari tahun sebelumnya yaitu 130/100.000 penduduk. Angka insiden tersebut jauh di bawah target pemerintah. Untuk tercapainya target maka diperlukan kebijakan-kebijakan pemerintah yang ditunjukkan untuk mengurangi jumlah kasus TB di Indonesia. Penyebaran bakteri *Mycobacterium Tuberculosis* yang begitu cepat dari satu orang ke orang lain, dari satu keluarga ke keluarga lain, bahkan dari suatu wilayah ke wilayah lain yang lebih luas. Penyebaran bakteri yang cepat dan luas mengakibatkan tingginya jumlah kasus TB. Untuk mengatasi hal tersebut diperlukan suatu model regresi yang memasukkan keterkaitan lokasi ke dalam model. Penelitian ini bertujuan untuk memodelkan jumlah kasus tuberkulosis di Indonesia tahun 2021 dengan faktor-faktor yang mempengaruhinya. Regresi spasial merupakan metode yang digunakan untuk mengetahui pengaruh variabel dependen dan variabel independen dengan mengamati pengaruh efek lokasi. Efek lokasi diukur menggunakan matriks pembobot *queen contiguity* modifikasi dengan jumlah kasus TB per 100.000 penduduk di 34 provinsi se-Indonesia. Berdasarkan hasil uji *lagrange multiplier* diperoleh adanya dependensi spasial *lag* dan *error* sehingga dilakukan pemodelan *Spatial Autoregressive Model (SAR)* dan *Spatial Error Model (SEM)*. Pada model diperoleh tiga variabel independen yang berpengaruh signifikan yaitu kepadatan penduduk (X_1), persentase rumah tangga dengan akses sanitasi layak (X_2), persentase tempat pengolahan pangan yang memenuhi syarat sesuai standar (X_3) dan didapatkan model terbaik yaitu *Spatial Autoregressive Model (SAR)*.

Kata kunci: *Lagrange Multiplier, Queen Contiguity, Regresi Spasial, Spatial Autoregressive Model (SAR), Spatial Error Model (SEM), Tuberkulosis.*

1. PENDAHULUAN

Tuberkulosis (TB) merupakan penyakit menular pada manusia yang terletak pada bagian bawah pernafasan yang disebabkan oleh bakteri *Mycobacterium Tuberculosis*. Penyebaran TB yang begitu cepat dari satu orang ke orang lain melalui udara ketika penderita batuk, bersin, ataupun meludah. Secara global angka kejadian TB di dunia masih cukup tinggi. Mengacu pada *World Health Organization (WHO)* dalam global TB Report 2020, penderita TB di dunia sebanyak 10 juta orang dengan 5,8 juta pria, 3,2 juta wanita dan 1 juta anak-anak. Indonesia menempati peringkat ketiga jumlah kasus TB tertinggi setelah India dan China. Jumlah orang jatuh sakit akibat TB di Indonesia diperkirakan mencapai 845.000 kasus dengan angka kematian sebanyak 98.000 atau sebanding dengan 11 kematian/jam [11].

Penanganan untuk mengurangi jumlah kasus TB dimana WHO membuat target global yaitu mengurangi kejadian TB sebanyak 80%, kematian TB sebanyak 90% dan untuk mentiadakan beban biaya bencana TB bagi keluarga yang terkena dampak pada tahun 2030. Indonesia berkomitmen mencapai eliminasi TB yaitu menurunkan jumlah kasus TB menjadi 65/100.000 penduduk [3].

* Program Studi Matematika, FMIPA-Universitas Mataram
Email: merysolihah00@gmail.com

Pada tahun 2021 jumlah kasus TB yang ditemukan sebanyak 146 kasus per 100.000 penduduk meningkat dari tahun sebelumnya yaitu 130 kasus per 100.000 penduduk. Jumlah kasus tersebut masih jauh dari target pemerintah. Agar tercapainya target maka diperlukan kebijakan-kebijakan pemerintah yang ditunjukkan untuk mengurangi jumlah kasus TB di Indonesia. Salah satu upaya yang dapat dilakukan yaitu dengan menganalisis faktor-faktor yang mempengaruhi jumlah kasus TB menggunakan analisis regresi linier. Regresi linier adalah suatu hubungan antara variabel dependen dengan satu variabel independen atau lebih [7].

Penularan bakteri *Mycobacterium Tuberculosis* yang begitu pesat dari satu orang ke orang lain, dari satu keluarga ke keluarga lain ataupun dari suatu lokasi ke lokasi lain yang lebih besar dapat mengakibatkan peningkatan jumlah kasus TB. Tobler [1] mengungkapkan bahwa segala sesuatu saling berhubungan satu sama lainnya, tetapi sesuatu yang dekat lebih mempunyai pengaruh daripada sesuatu yang jauh. Jika dihubungkan dengan pernyataan Tobler, hubungan antar lokasi dapat dipertimbangkan sebagai sebab interfensi jumlah kasus TB di Indonesia. Sehingga perlu ditinjau lebih jauh variabel-variabel dan hubungan lokasi yang berpengaruh terhadap jumlah kasus TB dari efek spasial.

Suatu model regresi yang memasukkan keterkaitan lokasi ke dalam model, salah satunya yaitu regresi spasial. Regresi spasial merupakan metode yang digunakan untuk mengetahui pengaruh variabel independen terhadap variabel dependen dengan mengamati pengaruh efek lokasi. Berdasarkan tipe data, pemodelan spasial dibedakan menjadi pemodelan pendekatan titik dan area. Menurut Azkiyah [5] data titik menunjukkan lokasi yang berupa titik, misalnya titik pada *longitude* (garis bujur) dan *latitude* (garis lintang) dan data area menunjukkan lokasi yang berupa luasan, seperti provinsi, kabupaten, kota, dan sebagainya. Berdasarkan latar belakang tersebut, peneliti melakukan pemodelan jumlah kasus TB menggunakan regresi spasial berbasis area sehingga dapat mengetahui provinsi yang mana saja harus lebih diperhatikan penanganan kasus TB serta mengetahui faktor-faktor yang berpengaruh signifikan terhadap TB sebagai solusi untuk mengurangi jumlah kasus TB di Indonesia.

2. METODELOGI

Regresi Linier Berganda

Model regresi linier berganda yang menghubungkan variabel dependen dengan p variabel independen sebagai berikut [7]:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_j X_{ij} + \dots + \beta_p X_{ip} + \epsilon_i \quad (2.1)$$

i menunjukkan unit pengamatan untuk variabel dependen Y dan p menunjukkan jumlah variabel independen. n merupakan ukuran sampel ($i = 1, 2, \dots, n$). Terdapat $(p + 1)$ parameter β_j , $j = (0, 1, \dots, p)$ diestimasi pada model yang memasukkan intercept β_0 . Terdapat asumsi $n > (p + 1)$.

Asumsi Non Multikolinieritas

Multikolinieritas yaitu terdapat hubungan linier antar variabel independen dengan melihat nilai *Variance Inflation Factor* (VIF). VIF dihitung dari korelasi matriks dari variabel independen. Jika nilai R_j^2 mendekati 1 maka nilai VIF akan besar. $VIF > 10$ maka dinyatakan terjadi multikolinieritas [7].

$$VIF = \frac{1}{1-R_j^2} ; j = 1,2, \dots, p \quad (2.2)$$

dimana:

p = banyaknya variabel independen

R_j^2 = koefisien determinasi antara variabel independen ke j dengan variabel independen lainnya.

Uji Model Regresi

Uji Serentak

Pengujian kecocokan model regresi linier yaitu dengan menggunakan analisis varians/*Analysis of Varians* (ANOVA) [10].

Hipotesis:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_1: \text{minimal ada satu } \beta_j \neq 0, j = 1,2, \dots, p$$

Statistik uji:

$$F_{hitung} = \frac{MSR}{MSE} \quad (2.3)$$
$$MSR = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{p} \quad MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - (p + 1)}$$

dimana

p = jumlah variabel independen

n = jumlah pengamatan

Kriteria pengambilan keputusan yaitu tolak H_0 jika $F_{hitung} > F_{\alpha;v_1,v_2}$ dimana $v_1 = p$ dan $v_2 = n - p - 1$ (dengan $\alpha = 5\%$).

Uji Parsial

Uji parsial atau uji t digunakan untuk mengetahui apakah model regresi variabel independen secara parsial berpengaruh signifikan terhadap variabel dependen atau untuk memperlihatkan seberapa jauh pengaruh masing-masing variabel independen dalam menerangkan variansi variabel dependen [9].

Hipotesis:

$$H_0: \beta_j = 0$$

$$H_1: \beta_j \neq 0, j = 1,2, \dots, p$$

Statistik uji:

$$t_{hit} = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \quad (2.4)$$

Kriteria pengambilan keputusan yaitu tolak H_0 jika $|t_{hit}| > t_{n-p;\alpha/2}$ dimana $df = n - p - 1$ (dengan $\alpha = 5\%$)

Uji Asumsi Regresi Linier

1. Uji Normalitas

Untuk menguji kenormalan residual menggunakan metode *Kolmogorov-Smirnov*[8].

Hipotesis:

H_0 : residual berdistribusi normal

H_1 : residual tidak berdistribusi normal

Statistik uji:

$$D = \sup|S(\epsilon) - F(\epsilon)| \quad (2.5)$$

dimana

D = deviasi maksimum

$S(\epsilon)$ = probabilitas kumulatif dari data pengamatan/sampel

$F(\epsilon)$ = probabilitas kumulatif distribusi normal.

Daerah keputusan H_0 ditolak jika $D > D_{n,\alpha}$ dan $\alpha = 0,05$

2. Uji Homoskedastisitas

Asumsi homoskedastisitas merupakan kondisi dimana varians untuk setiap nilai *error* konstan pada semua variabel independen[8].

Hipotesis:

H_0 : $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2$ (homoskedastisitas)

H_1 : minimal ada satu $\sigma_i^2 \neq \sigma_j^2$ (heteroskedastisitas)

Statistika uji:

$$\chi_{hitung}^2(BP) = n * R_{aux}^2 \quad (2.6)$$

dimana:

R_{aux}^2 = koefisien determinasi dari *error* kuadrat (ϵ^2)

n = jumlah pengamatan

Keputusan tolak H_0 jika $\chi_{hitung}^2 > \chi_{tabel}^2$ dengan $\alpha = 0,05$.

3. Uji Autokorelasi

Menurut Sari [8] asumsi autokorelasi merupakan asumsi untuk mengidentifikasi bawasannya ada satu atau lebih variabel independen yang mempengaruhi variabel dependen dan tidak dimasukkan dalam model regresi.

Hipotesis:

H_0 : tidak ada autokorelasi

H_1 : terdapat autokorelasi

Menggunakan statistik uji *Durbin Watson* yang dinotasikan sebagai:

$$d = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2} \quad (2.7)$$

dimana:

d = nilai Durbin Watson

e_i = *error* pada i

e_{i-1} = *error* pada satu lag sebelumnya

Keputusan:

Tolak H_0 jika $d < dl$ atau $d > 4 - du$ atau $p\text{-value} < \alpha$

Uji Efek Spasial

1. Uji Indeks Moran

Adanya hubungan pengamatan lokasi i dengan pengamatan lain di lokasi j dimana $j \neq i$ dapat diketahui dengan uji autokorelasi spasial menggunakan Indeks Moran (I). Menurut yasin *et al* [12] autokorelasi spasial adalah taksiran dari korelasi antar nilai amatan yang berkaitan dengan lokasi pada variabel yang sama.

Hipotesis:

H_0 : Tidak ada autokorelasi spasial antar lokasi

H_1 : Ada autokorelasi spasial antar lokasi

Statistik Uji:

$$Z_{hitung} = \frac{I - E(I)}{\sqrt{Var(I)}} \quad (2.8)$$

dengan:

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} c_{ij}}{\sum_{i=1}^n \sum_{j=1}^n w_{ij} \sum_{i=1}^n (x_i - \bar{x})^2}, \quad c_{ij} = (x_i - \bar{x})(x_j - \bar{x})$$

$$E(I) = I_0 = -\frac{1}{n-1} \quad Var(I) = \frac{n^2 S_1 - n S_2 + 3 S_0^2}{(n^2 - 1) S_0^2} - [E(I)]^2$$

$$S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij}; S_1 = \frac{1}{2} \sum_{i \neq j} (w_{ij} + w_{ji})^2; S_2 = \sum_{i \neq j} (w_{i0} + w_{0i})^2; w_{i0} = \sum_{j=1}^n w_{ij}; \sum_{j=1}^n w_{0i} = \sum_{j=1}^n w_{ji}$$

Keputusan: H_0 ditolak jika $|Z_{hitung}| > Z_{\alpha/2}$ dengan $\alpha = 0,05$

2. Uji Lagrange Multiplier

Lagrange Multiplier digunakan sebagai dasar untuk memilih model regresi spasial [6].

LM uji untuk SAR

Hipotesis:

$H_0 : \rho = 0$ (tidak ada lag dependensi spasial)

$H_1 : \rho \neq 0$ (ada lag dependensi spasial)

Statistik uji yang digunakan adalah:

$$LM_{lag} = \frac{\left(\frac{\epsilon^T W y}{s^2}\right)^2}{\frac{(W X \beta)^T M (W X \beta)}{s^2} + C} \quad (2.9)$$

dimana

$$M = I - X(X^T X)^{-1} X^T \quad C = tr((W^T + W)W) \quad s^2 = \frac{(\epsilon^T \epsilon)}{n}$$

dengan

X = matriks variabel independen berukuran $n \times k$

W = matriks pembobot spasial berukuran $n \times n$

ϵ = vektor *error* dari model regresi

y = vektor variabel dependen berukuran $n \times 1$

I = matriks identitas berukuran $n \times n$

Kriteria uji adalah tolak H_0 jika $LM_{lag} > \chi^2_{(1),\alpha}$ dengan $\alpha = 0,05$.

LM uji untuk SEM

Hipotesis:

$H_0 : \lambda = 0$ (tidak ada dependensi *error* spasial)

$H_1 : \lambda \neq 0$ (ada dependensi *error* spasial)

Statistik uji yang digunakan adalah:

$$LM_{error} = \frac{\left(\frac{\epsilon^T W \epsilon}{s^2}\right)^2}{C} \quad (2.10)$$

dimana:

$$C = tr(W^T + W)W$$

Kriteria uji adalah tolak H_0 jika $LM_{error} > \chi^2_{(1),\alpha}$ dengan $\alpha = 0,05$.

Spatial Autoregressive Model (SAR)

SAR disebut juga *Spatial Lag Model* (SLM) adalah salah satu model spasial dengan pendekatan lokasi yang memperhitungkan pengaruh spasial lag (ρ) hanya pada variabel dependen. Model ini merupakan kombinasi dari model regresi biasa dengan model regresi spasial lag pada variabel dependen [12]. Secara matematis model SAR dapat ditulis sebagai berikut:

$$\begin{aligned} \mathbf{y} &= \rho \mathbf{W} \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &\sim N(0, \sigma_{\boldsymbol{\varepsilon}}^2) \end{aligned} \quad (2.11)$$

Spatial Error Model (SEM)

SEM dapat digunakan ketika nilai *error* pada suatu lokasi berkorelasi dengan nilai *error* dengan lokasi sekitarnya atau dengan kata lain terdapat korelasi spasial antar *error* (λ). Pada model SEM, bentuk *error* pada lokasi i merupakan fungsi dari *error* pada lokasi j dimana j merupakan suatu lokasi yang terletak disekitar lokasi i [12]. Secara matematis model SEM dapat ditulis sebagai berikut:

$$\mathbf{y} = \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2.12)$$

dengan

$$\boldsymbol{\varepsilon} = \lambda \mathbf{W} \boldsymbol{\varepsilon} + \mathbf{v}$$

Pemilihan Model terbaik

Model dengan nilai AIC yang kecil adalah yang terbaik. Metode AIC dirumuskan sebagai berikut [1]:

$$AIC = -2 \log L(\hat{\theta}_k) + 2k \quad (2.13)$$

Koefisien determinasi R^2 menunjukkan ketepatan suatu model dinotasikan [4]:

$$R^2 = 1 - \frac{\sigma_{ML}^2}{\sigma_Y^2} \quad (2.14)$$

dengan

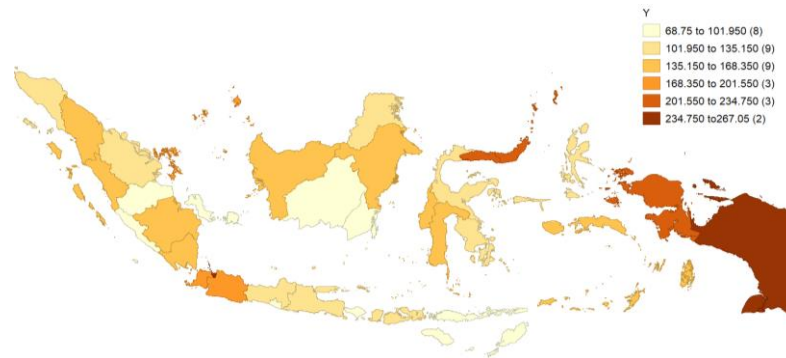
$$\sigma_Y^2 = \frac{\sum_{i=1}^n (y_i - \bar{Y})^2}{N - 1}$$

Nilai dari koefisien determinasi ini adalah $0 \leq R^2 \leq 1$. Semakin besar R^2 maka kepercayaan terhadap model semakin besar, model semakin tepat dalam menggambarkan fenomena dari variabel dependen.

Data

Penelitian ini menggunakan data sekunder yaitu setiap provinsi di Indonesia tahun 2021 terdiri dari 34 provinsi yang diperoleh dari hasil publikasi Profil Kesehatan Indonesia 2021, Kesehatan Indonesia 2021 dan Statistik Indonesia 2022 serta peta wilayah batas provinsi dukcapil desember 2019 diperoleh dari *website* Indonesia Geospasial. Variabel yang digunakan dalam penelitian ini adalah Jumlah kasus tuberkulosis per 100.000 penduduk (jiwa) (Y), yaitu kepadatan penduduk (X_1), persentase rumah tangga dengan akses sanitasi layak (X_2), persentase tempat pengolahan pangan yang memenuhi syarat sesuai standar (X_3), persentase pengobatan lengkap tuberkulosis (X_4), persentase rumah layak huni (X_5), persentase gizi buruk pada balita (X_6). Matriks pembobot (W) yang digunakan yaitu *queen contiguity* modifikasi dimana wilayah yang berdekatan atau bertetangga diberi bobot 1 dan yang tidak bertetangga diberi bobot 0 serta dilakukan modifikasi dimana wilayah yang berdekatan yang dipisahkan oleh lautan dianggap tetap bertetangga.

3. HASIL DAN PEMBAHASAN



Gambar 3.1 Peta persebaran jumlah kasus tuberkulosis tahun 2021

Eksistensi persebaran jumlah kasus tuberkulosis per 100.000 penduduk (Y) di Indonesia tahun 2021 dapat dilihat pada gambar 3.1 dimana semakin gelap warna menunjukkan semakin besar jumlah kasus TB, sebaliknya semakin terang warna menunjukkan semakin rendah jumlah kasus TB. Wilayah-wilayah yang berdekatan cenderung memiliki warna yang sama sehingga dapat dikelompokkan berdasarkan jumlahnya atau dapat dikatakan adanya keterkaitan suatu wilayah dengan wilayah sekitarnya.

Pemodelan Regresi Linier

Model diperoleh dengan pengolahan menggunakan *software* R Studio dengan model regresi linier sebagai berikut.

$$Y_i = 383,534114 + 0,010614X_{i1} - 1,862360X_{i2} - 0,904422X_{i3} - 0,822645X_{i4} - 0,038776X_{i5} + 8,258074X_{i6} + \epsilon_i$$

Uji Serentak

Berdasarkan kriteria pengambilan keputusan yaitu tolak H_0 jika $F_{hitung} > F_{\alpha;v_1,v_2}$ dimana $v_1 = p$ dan $v_2 = n - p - 1$. Pada hasil perhitungan menggunakan persamaan (2.3) diperoleh nilai uji F sebesar 3,753198 dengan $F_{0,05;6;27} = 2,38$. Dapat disimpulkan $F_{hitung} > F_{tabel}$ maka tolak

H_0 sehingga variabel-variabel independen secara bersama-sama cenderung berpengaruh terhadap variabel dependen.

Uji Parsial

Pengolahan data menggunakan R studio diperoleh output sebagai berikut

Tabel 3.1 Penduga parameter model regresi linier

Variabel	Koefisien	Standar error	t_{hit}	
<i>intercept</i>	83,534114	96,240792	3,9851513	*
X_1	0,010614	0,003024	3,5100329	*
X_2	-1,862360	01,006596	-1,850155	.
X_3	-0,904422	0,525216	-1,722002	.
X_4	-0,822645	1,81540	-0,696248	
X_5	-0,038776	0,719550	-0,053889	
X_6	8,258074	13,971651	0,5910593	

Kriteria pengambilan keputusan yaitu tolak H_0 jika $|t_{hit}| > t_{n-p;\alpha/2}$. Berdasarkan tabel 3.1 menunjukkan jika $|t_{hit}|$ lebih besar dari $t_{28;0,05/2} = 2,048$ (ditandai dengan " * ") dan lebih besar dari $t_{28;0,1/2} = 1,701$ (ditandai dengan ". ") maka variabel tersebut cenderung berpengaruh signifikan terhadap jumlah kasus tuberkulosis. Hasil yang diperoleh yaitu terdapat tiga variabel independen yang signifikan pada $\alpha = 0,1$ yaitu X_1, X_2 dan X_3 , terdapat satu variabel independen yang signifikan pada $\alpha = 0,05$ yaitu X_1 .

Uji Normalitas

Pengolahan data menggunakan R studio diperoleh output sebagai berikut

```
Lilliefors (Kolmogorov-Smirnov)
normality test

data:  Reg_OLS$residuals
D = 0.086741, p-value = 0.7441
```

Daerah keputusan H_0 ditolak jika $D > D_{n,\alpha}$ dengan $\alpha = 0,05$. Berdasarkan hasil pengolahan data diperoleh nilai D sebesar 0,086741 lebih kecil dari D_{tabel} sebesar 0,22424 maka gagal tolak H_0 , dengan demikian dapat disimpulkan bahwa residual regresi linier berdistribusi normal.

Uji Homoskedastisitas

Pengolahan data menggunakan R studio diperoleh output sebagai berikut

```
studentized Breusch-Pagan test

data:  Reg_OLS
BP = 7.3923, df = 6, p-value = 0.2861
```


Keputusan tolak H_0 jika $\chi^2_{hitung}(BP) > \chi^2_{tabel}$ dengan $\alpha = 0,05$. Berdasarkan hasil pengolahan data diperoleh nilai $\chi^2_{hitung} = 7,3923$ lebih kecil dari $\chi^2_{tabel} = 12,592$ maka gagal tolak H_0 dengan demikian tidak terjadi heteroskedastisitas.

Uji Autokorelasi

Pengolahan data menggunakan R studio diperoleh output sebagai berikut

```
Durbin-Watson test

data:  Reg_OLS
DW = 1.4988, p-value = 0.03803
alternative hypothesis: true autocorrelation is greater than 0
```

Berdasarkan tabel DW diperoleh *durbin Upper* (du) = 1,8913 dan *Durbin Lower* (dl) = 1,0794 maka nilai d berada di antara nilai dl dan du sehingga tidak ada kesimpulan atau tidak pasti. Namun dapat dilihat nilai $p\text{-value}$ $0,03803 < \alpha = 5\%$ maka tolak H_0 yang berarti terdapat autokorelasi. Karena asumsi regresi linier non autokorelasi tidak terpenuhi maka regresi linier bukan model yang tepat untuk memodelkan jumlah kasus tuberkulosis di Indonesia tahun 2021.

Apabila saat menganalisis model regresi linier menggunakan data spasial, maka bisa menyebabkan kesimpulan tidak tepat karena asumsi *error* saling bebas (autokorelasi) penyebabnya yaitu data spasial mempunyai sifat dimana errornya saling berkorelasi (Arif et al., 2019) Oleh karena itu diperlukan pertimbangan aspek spasial dengan mengecek efek spasial menggunakan Indeks moran dan Lagrange Multiplier.

Identifikasi Efek Spasial

Indeks Moran

Pengolahan data menggunakan R studio diperoleh output sebagai berikut

Tabel 3.2 Uji Indeks Moran

Indeks Moran (I)	Expected Value $E(I)$	Varians
0,51340467	-0,03030303	0,02803725

Berdasarkan nilai pada tabel 3.2 dengan menggunakan persamaan (2.8) diperoleh $Z_{hitung} = 3,2471$ lebih besar dari $Z_{tabel} = Z_{0,975} = 1,96$ sehingga dapat disimpulkan tolak H_0 maka terdapat autokorelasi spasial antar wilayah yaitu jumlah kasus TB di provinsi di Indonesia memiliki hubungan dengan jumlah kasus TB di provinsi lainnya yang bertetanga.

Lagrange Multiplier

Pengolahan data menggunakan R studio diperoleh output sebagai berikut

Tabel 3.3 Hasil uji LM

LM	Statistika uji LM
LM_{lag}	8,3689
LM_{error}	6,3698

Berdasarkan tabel 3.3 baik statistik uji LM lag maupun statistic uji LM *error* lebih besar dari nilai $\chi^2_{tabel} = \chi^2_{(1),0,05} = 3,840$ sehingga tolak H_0 maka terdapat lag dependensi spasial dan terdapat dependensi *error* spasial. Oleh karena itu model regresi spasial yang dapat digunakan yaitu *Spatial Autoregressive Model* (SAR) dan *Spatial Error Model* (SEM).

Spatial Autoregressive Model (SAR)

Berikut adalah hasil estimasi model SAR menggunakan software R studio.

Tabel 3.4 Hasil estimasi model SAR

Parameter	Koefisien	Standar Error	Wald	
<i>intercept</i>	308,2694276	74,1725373	17,27327	*
$\hat{\beta}_1$	0,0091978	0,0023118	15,82951	*
$\hat{\beta}_2$	-1,6874012	0,74455668	5,136191	*
$\hat{\beta}_3$	-0,7395656	0,3926975	3,546804	.
$\hat{\beta}_4$	-0,9593230	0,8728842	1,20786	
$\hat{\beta}_5$	0,1249587	0,5316003	0,055254	
$\hat{\beta}_6$	-0,3257098	10,6001230	0,000944	
$\hat{\rho}$	0,41908	0,11679	12,87605	*

Kriteria uji yaitu tolak H_0 yang berarti parameter signifikan jika nilai *Wald* lebih besar dari $\chi^2_{(1),0,05} = 3,8415$ dan $\chi^2_{(1),0,1} = 2,706$. Berdasarkan hasil tabel 3.4 dapat dilihat nilai *Wald* pada parameter-parameter yang signifikan pada taraf 0,05 ditandai dengan tanda (*) yaitu X_1 dan X_2 dan signifikan pada taraf 0,1 ditandai dengan (.) yaitu X_3 . Berdasarkan persamaan (2.11) diperoleh model SAR sebagai berikut:

$$y_i = 0,41908 \sum_{j=1, i \neq j}^{34} W_{ij} Y_j + 308,2694276 + 0,0091978 X_{1i} - 1,6874012 X_{2i} - 0,7395656 X_{3i} + \varepsilon_i \quad (3.1)$$

Uji Normalitas Error SAR

Pengolahan data menggunakan R studio diperoleh output sebagai berikut

```
Lilliefors (Kolmogorov-Smirnov) normality
test

data: SAR$residuals
D = 0.093092, p-value = 0.6408
```

Daerah keputusan H_0 ditolak jika $D > D_{n,\alpha}$ dengan $\alpha = 0,05$. Dapat dilihat nilai $D_{hitung} = 0,093092 < D_{tabel} = 0,22424$ maka gagal tolak H_0 , dengan demikian *error* model SAR berdistribusi normal.

Uji homoskedastisitas Error SAR

Pengolahan data menggunakan R studio diperoleh output sebagai berikut

```
studentized Breusch-Pagan test

data:
BP = 7.4064, df = 6, p-value = 0.2849
```

Keputusan tolak H_0 jika χ^2_{hitung} (BP) $>$ χ^2_{tabel} dengan $\alpha = 0,05$. Berdasarkan hasil pengolahan data diperoleh nilai $\chi^2_{hitung} = 7,4064$ lebih kecil dari $\chi^2_{tabel} = 12,592$ maka gagal tolak H_0 , dengan demikian tidak terjadi heteroskedastisitas. Sehingga dapat disimpulkan bahwa asumsi model terpenuhi karena *error* model SAR berdistribusi normal dan memenuhi asumsi identik (homoskedastisitas).

Spatial Error Model (SEM)

Berikut adalah hasil pengolahan data model SEM menggunakan *software* R Studio.

Tabel 3.5 Hasil estimasi model SEM

Parameter	Koefisien	Standar Error	Wald	
<i>intercept</i>	354,9945386	65,6018008	26,2827474	*
$\hat{\beta}_1$	0,0099864	0,0024434	16,70431903	*
$\hat{\beta}_2$	-1,6463308	0,6980163	5,562923372	*
$\hat{\beta}_3$	-0,8468653	0,3960921	4,571264124	*
$\hat{\beta}_4$	-1,0184569	0,8495499	1,43716795	
$\hat{\beta}_5$	0,3532791	0,6111403	0,334159679	
$\hat{\beta}_6$	3,5103118	10,5455583	0,110803182	
$\hat{\lambda}$	0,50333	0,12159	17,13601825	*

Kriteria uji yaitu tolak H_0 yang berarti parameter signifikan jika nilai *Wald* lebih besar dari $\chi^2_{(1),0,05} = 3,840$. Berdasarkan hasil tabel 3.5 dapat dilihat nilai *Wald* pada parameter-parameter yang signifikan pada taraf 0,05 ditandai dengan tanda (*) yaitu X_1, X_2 dan X_3 . Berdasarkan persamaan (2.12) diperoleh model SEM dengan parameter yang signifikan sebagai berikut:

$$y_i = 354,9945386 + 0,0099864X_{1i} - 1,6463308X_{2i} - 0,8468653X_{3i} + 0,50333 \sum_{j=1, i \neq j}^{34} W_{ij}\epsilon_j + v_i \tag{3.2}$$

Uji Normalitas Error SEM

Pengolahan data menggunakan R studio diperoleh output sebagai berikut

```
Lilliefors (Kolmogorov-Smirnov)
normality test

data: SEM$residuals
D = 0.086571, p-value = 0.7468
```

Daerah keputusan H_0 ditolak jika $D > D_{n,\alpha}$ dengan $\alpha = 0,05$. Berdasarkan hasil pengolahan data diperoleh nilai D sebesar 0,086571 lebih kecil dari D_{tabel} sebesar 0,22424 maka gagal tolak H_0 , dengan demikian dapat disimpulkan bahwa *error* model SAR berdistribusi normal.

Uji Homoskedastisitas *Error SEM*

Pengolahan data menggunakan R studio diperoleh output sebagai berikut

```
studentized Breusch-Pagan test  
  
data:  
BP = 4.0013, df = 6, p-value = 0.6765
```

Keputusan tolak H_0 jika $\chi_{hitung}^2 > \chi_{tabel}^2$ dengan $\alpha = 0,05$. Berdasarkan hasil pengolahan data diperoleh nilai $\chi_{hitung}^2 = 4,0013$ lebih kecil dari $\chi_{tabel}^2 = 12,592$ maka gagal tolak H_0 , dengan demikian tidak terjadi heteroskedastisitas.

Sehingga dapat disimpulkan bahwa asumsi model terpenuhi karena *error* model SEM berdistribusi normal dan memenuhi asumsi identik (homogenitas).

Pemilihan Model terbaik

Diperoleh nilai AIC, R^2 antara SAR dan SEM sebagai berikut

Tabel 3.6 Nilai AIC model SAR dan SEM

Model	AIC	R^2
SAR	347,3878	0,6253
SEM	347,9305	0,6368

Pada tabel 3.6 menunjukkan nilai AIC model SAR lebih kecil dari nilai AIC model SEM dan R^2 model SEM lebih besar daripada R^2 model SAR. Pada model regresi nilai R^2 merupakan nilai yang semakin besar jika ditambahkan variabel independen walaupun variabel independen memberikan sedikit tambahan atau tidak ada perbaikan (Steenbeek & Ruiters, 2021). Berdasarkan pernyataan tersebut nilai R^2 pada model spasial yang berbeda tidak dapat digunakan untuk memilih model karena perbandingan model namun dapat digunakan sebagai ukuran ketepatan suatu model. Berdasarkan nilai AIC dapat disimpulkan SAR merupakan model terbaik dalam memodelkan jumlah kasus tuberkulosis di Indonesia Tahun 2021 dengan ketepatan suatu model sebesar 62% yang berarti 62% sebaran variabel dependen dapat dijelaskan oleh variabel independen dan pengaruh wilayah, sisanya 38% tidak dapat dijelaskan oleh variabel independen atau dapat dijelaskan oleh variabel diluar variabel independen.

4. KESIMPULAN

Berdasarkan hasil analisis dan pembahasan sebelumnya diperoleh model terbaik dalam memodelkan jumlah kasus tuberkulosis di Indonesia tahun 2021 yaitu *Spatial Autoregressive Model* (SAR) dengan nilai AIC 347,3878 dan R^2 0,6253 serta variabel independen yang berpengaruh signifikan yaitu kepadatan penduduk (X_1), persentase rumah tangga dengan akses sanitasi layak (X_2) dan persentase tempat pengolahan pangan yang memenuhi syarat sesuai standar (X_3).

REFERENSI

- [1] Anselin, L. (1988). *Spatial Econometrics: Methods and Models* (Vol. 4). Springer Netherlands. <https://doi.org/10.1007/978-94-015-7799-1>
- [2] Arif, A., Tiro, M. A., & Nusrang, M. (2019). Perbandingan Matriks Pembobot Spasial Optimum Dalam Spatial Error Model (SEM) (Kasus : Indeks Pembangunan Manusia Kabupaten/Kota di Provinsi Sulawesi Selatan Tahun 2015). *Variansi: Journal of Statistics and Its Application on Teaching and Research*, 1(3).
- [3] Kemenkes. 2021. Jadikan Penerus Bangsa Bebas TBC, Dimulai dari Diri Sendiri dan Keluarga. <https://www.kemkes.go.id>. [22 November 2022]
- [4] Kosfeld, R. 2020. Spatial Econometrics. University of Kassel. www.uni-kassel.de. [20 November 2022]
- [5] Lutfiani, N., & Scolastika Mariani, dan. (2019). Pemodelan Geographically Weighted Regression (GWR) dengan Fungsi Pembobot Kernel Gaussian dan Bi-square. In *UNNES Journal of Mathematics* (Vol. 5, Issue 1). <http://journal.unnes.ac.id/sju/index.php/ujmUJM8>
- [6] Rahayu, S. P., & Pratiwi, N. (2022). Aplikasi Metode Regresi Spasial Data Panel Pada Penerimaan Daerah Sektor Pariwisata di Kabupaten/Kota di Daerah Istimewa Yogyakarta. *Jurnal Statistika Industri dan Komputasi*, 7(1), 13–23.
- [7] Rawlings, J. O., Pantula, S. G., & Springer, D. A. D. (n.d.). *Applied Regression Analysis: A Research Tool, Second Edition*.
- [8] Sari, I. M. (2020). Analisis Regresi Linear Berganda Pada Sisa Hasil Usaha (SHU) Koperasi (Studi Pada Koperasi Kredit Bunga Tanjung Lampung Selatan).
- [9] Taek, Y., Dwi Bekti, R., Suryowati, K., & Statistika, J. (2023). Penerapan Model Geographically Weighted Regression (Gwr) Menggunakan Fungsi Pembobot Adaptive Kernel Gaussian Dan *Adaptive Kernel Bisquare* Pada Tingkat Pengangguran Terbuka Di Pulau Papua. *Jurnal Statistika Industri Dan Komputasi*, 08(2), 84–101. <https://papua.bps.go.id>
- [10] Weisberg, S. (n.d.). *Applied Linear Regression Third Edition*. www.copyright.com.
- [11] World Health Organization. (2020). *Global tuberculosis report 2020*. World Health Organization.
- [12] Yasin Budi Warsito Arief Rachman Hakim, H. (n.d.). *Regresi Spasial (Aplikasi dengan R)*.