Identifying Factors that Influence Life Expectancy in Central Java Using Spatial Regression Models

Prizka Rismawati Arum(1), Rahmad Putra Gautama(2), Indah Fitriani(3), Fellya Naza Nurvahyani(4).

[1,2,3,4] Departements of Statistics, Universitas Muhammadiyah Semarang, 20273, Semarang, Indonesia

e-mail: prizka.rismawatiarum@unimus.ac.id(1), rahmadgautama15@gmail.com(2), indahfitriyani110@gmail.com(3), fellya2181@gmail.com(4)

ABSTRAK

Angka Harapan Hidup adalah rata-rata harapan hidup penduduk dalam beberapa tahun, dengan asumsi angka kematian tetap konstan seiring bertambahnya umur. Fungsinya sebagai alat pengukur keberhasilan pembangunan kesehatan penduduk di perkotaan dan kesejahteraan umum, terutama dalam aspek kesehatan. Tinggi rendahnya angka harapan hidup dipengaruhi oleh beberapa indikator, seperti kondisi sosial ekonomi, lingkungan, dan kesehatan. Penelitian ini bertujuan mengidentifikasi komponen-komponen penting yang memengaruhi angka harapan hidup di 35 kabupaten dan kota di Provinsi Jawa Tengah melalui pendekatan proses dalam menganalisis metode regresi spasial. Selain itu, penelitian ini mencari persamaan regresi spasial terbaik untuk pemodelan angka harapan hidup di Provinsi Jawa Tengah. Regresi spasial adalah metode pengembangan regresi linier yang tergolong dalam elemen titik model. Menggunakan dua variabel independen terpilih dari tujuh variabel independen, penelitian ini mempelajari persamaan regresi spasial dengan pendekatan wilayah SAR, SEM, dan SARMA. Hasilnya menunjukkan bahwa model SAR terpilih dengan nilai p-value 0,02183 yang sesuai untuk mengidentifikasi ketergantungan efek spasial terhadap angka harapan hidup di Jawa Tengah. Tingkat Pengangguran Terbuka (X4) dan Persentase Penduduk Miskin (X6) adalah faktor signifikan yang memengaruhi angka harapan hidup di Jawa Tengah secara spasial.

Kata Kunci: Angka Harapan Hidup, Regresi Spasial, Jawa Tengah, Log Range Multiplier (SAR)

ABSTRACT

Life Expectancy is an average calculated over several years, assuming that mortality remains constant as age increases. It serves as a metric to gauge the success of population health development at the urban level and overall well-being, particularly in terms of health. Various indicators, including socioeconomic conditions, environmental factors, and health indicators, influence the highs and lows of life expectancy. This study in Central Java Province's 35 districts and cities aims to identify crucial components impacting life expectancy through a process-oriented spatial regression analysis. Additionally, the research endeavors to determine the optimal spatial regression equation for modeling life expectancy in the province. Spatial regression, a linear regression development method falling under the point element model, is employed. Utilizing two independent variables selected from seven, the study explores spatial regression equations using SAR, SEM, and SARMA area approaches. Data sourced from BPS in 2020 reveals that the SAR model, with a p-value of 0.02183, is apt for identifying spatial effects on Central Java's life expectancy. The Open Unemployment Rate (X_4) and the Percentage of Poor Population (X_6) emerge as significant spatial factors influencing life expectancy in Central Java.

Keywords: *Life Expectancy, Spatial Regression, Central Java, Lagrange Multiplier Log (SAR)*

INTRODUCTION

The Life Expectancy Rate is an estimate of the average lifespan of a population over several years, assuming that mortality does not change with age. In developing countries like Indonesia, the Life Expectancy Rate is considered low due to factors such as inadequate healthcare facilities, making the population susceptible to clinical disorders. As a result, mortality rates in developing countries are much higher compared to those in developed countries, leading to a lower life expectancy in these nations. The Life Expectancy Rate can be evaluated by examining the effectiveness of healthcare development for the population, measuring urbanization levels, and promoting the general well-being of the population, particularly in terms of health status. Several indicators, including health indicators, environmental conditions, and socio-economic factors, play a crucial role, with the most important indicators reflecting the level of life expectancy seen from the development of the health sector in a region, which has a significant impact on determining the standard of life expectancy [1].

According to data released by the Population Reference Bureau, Indonesia's Life Expectancy (LE) has been consistently below the global average each year, ranging from 69 to 71 years in 2018-2020 and increasing slightly to 72-73 years in subsequent year [21] s. In the period of 2018-2020, Indonesia held the 7th position in Life Expectancy among ASEAN countries. Based on statistics released by the Central Statistics Agency (BPS), Indonesia's Life Expectancy reached 71.6 years in 2021, reflecting a slight increase of 0.1 years compared to the previous year when it was 71.5 years in 2020. It's noteworthy that the Life Expectancy for females in Indonesia is higher compared to males [21].

Java Island is one of the large island clusters in Indonesia. According to the 2022 data from BPS, Indonesia has a population of 275.77 million people, with 56.05% of the population residing on Java Island. The island's high population is attributed to the various attractions it possesses, and it is also the economic center of Indonesia. The LE in Central Java Province ranks third, with a 2021 LE score of 74.5 years, indicating a 0.1-year increase from 2020. However, despite this improvement, Central Java couldn't maintain its position as the second-highest life expectancy province as it had in the previous year. In 2021, the average Life Expectancy in Central Java Province showed that residents in Sukoharjo Regency had the highest LE at 77.73 years, while residents in Brebes Regency had the lowest LE at 69.54 years [2].

From what has been explained, this research employs spatial regression analysis to examine the significant factors influencing Life Expectancy (LE) in Central Java and obtain the best model for modeling Life Expectancy in the Central Java Province. Linear regression analysis is a type of statistical analysis used to explain the correlation between independent and dependent variables. The method used to analyze relationships between different variables is called spatial regression. This method considers the spatial influence at different locations, making it the focus of the research. This issue emphasizes the possibility that spatial dependence in data, spatial dependence, and spatial variability can impact spatial data. [3]. Spatial modeling of this kind uses line and area methods. Spatial Autoregressive (SAR) is based on spatial lag effects, Spatial Error Model (SEM) is based on spatial error effects, and Spatial Autoregressive Moving Average (SARMA) are three crucial spatial methods for area analysis based on a combination of lag and spatial error effects [4].

The previous research, cited from the Journal of Data Analysis and written by Evi Ramadhani, Nani Salwa, and Madina Suha Mazaya (2022), regarding the analysis of factors using spatial regression, concludes that the factors influencing Life Expectancy in Sumatra in 2018 can be analyzed using the best area approach, namely the Spatial Error Model (SEM) [5]. As a reference and another research source, a study by Fatkhurokhman Fauzi (2016) on the Best Spatial Regression Model for the Human Development Index in Central Java Province yielded the conclusion that Life Expectancy reflects the level and degree of health of a community [6].

This research aims to utilize spatial regression analysis to examine the significant factors influencing Life Expectancy (LE) across various regencies and cities in the Central Java Province. Additionally, the study strives to identify the best spatial regression model for accurately simulating LE, forecasting LE values in the future, and pinpointing the variables that influence LE in the province of Central Java. It is expected that insights into the LE model for each regency and city can assist communities and governments in promoting LE values, especially in less common areas.

METHODOLOGY

Data

The data used in this study is secondary data sourced from the Central Statistics Agency (Badan Pusat Statistik) of Central Java Province, consisting of 35 regencies/cities. The variables employed include Life Expectancy (Y), Proportion of Malnourished and Undernourished Toddlers (X_1), Proportion of Households with Sanitation Access (X_2), Proportion of Households with Access to Clean Water (X_3), Open Unemployment Rate (X_4), Real Per Capita Expenditure (X_5), Percentage of Poor Population (X_6), and Labor Force Participation Rate (X_7). These variables are observed across all 35 regencies/cities in Central Java Province.

Data Analysis Procedure

The following are the analysis steps conducted in the study:

1. Creating data based on the research variables used.
2. Performing a descriptive exploration of the research data.
3. Using Pearson correlation analysis to select independent variables.

$$r = \frac{n\sum_{i=1}^{n} X_i Y_i - (\sum_{i=1}^{n} X_i) - (\sum_{i=1}^{n} Y_i)}{\sqrt{[n\sum_{i=1}^{n} {X_i}^2 - (\sum_{i=1}^{n} X_i)^2][n\sum_{i=1}^{n} {Y_i}^2 - (\sum_{i=1}^{n} Y_i)^2]}}$$

4. Conducting multiple linear regression analysis on life expectancy (LE) in Central Java, involving 35 regencies and cities and 7 independent variables.

$$Y_i = \beta_0 + \sum_{k=1}^{p} \beta_k X_{ki} + \varepsilon_i$$

5. Creating a spatial weighting matrix (***W***) using the *Queen Contiguity* method.

$$W = \begin{bmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{n1} & \cdots & w_{nn} \end{bmatrix}$$

Formation of the result matrix used in the model, where:

$$W_{ij}^{*} = \frac{W_{ij}}{\sum_{i=1}^{n} w_{ij}}$$

6. Testing spatial dependence and heterogeneity using:
   a. *Moran's I* Test
      The spatial autocorrelation of residuals is tested using Moran's I test. Spatial autocorrelation is employed to assess whether there is correlation among members of a series and observations decomposed according to time and space.
   b. Spatial Heterogeneity Test
      Breusch-Pagan Test, testing spatial heterogeneity.
7. Determining the spatial regression model using the *Lagrange Multiplier (LM)* test.

$$LM_{SARMA} = E^{-1}\left\{\left(R_y\right)^2 T - 2R_y R_\varepsilon T + (R_\varepsilon)^2 (D + T)\right\} \sim \chi^2_{(k)}$$

8. Calculating and testing the parameters of the regression analysis model.
9. Analyzing the spatial regression model and drawing conclusions.

RESULTS AND DISCUSSION

Figure 1 illustrates the distribution map of Life Expectancy (LE) in Central Java, divided into five interval groups. This division aims to provide a more detailed overview of LE variations in the region. Each interval group encompasses a specific range of LE values representing categories within each section of the map. Visual observations on the map provide information about LE levels in Central Java, where an increasing color intensity indicates higher LE values. The use of different color scales offers clear visual cues about LE variations across the region. The transition of colors from light to dark reflects an increase in LE values. The presentation of information through this map is designed to enhance understanding of the geographic distribution of LE in Central Java. This approach provides a comprehensive insight into the health and life expectancy disparities in the region. Therefore, this map can serve as a valuable tool in the context of public health analysis and decision-making for related policy considerations.

Figure 1. LE data distribution map for Central Java Province

In Figure 1, the map illustrates the distribution of Life Expectancy (LE) in Central Java divided into five interval groups. The increasing intensity of color on the map indicates higher LE values. Summary statistics can provide information about the central tendency, dispersion, and trend patterns of the data. The average LE value in regencies/cities in Central Java in 2021 is 75.012 years, with a median value of 74.84 years. A higher mean value compared to the median suggests that the distribution of LE data in Central Java in 2018 tends to be right-skewed (positive skewness). The highest LE achievement, with a value of 77.73 years, is in Sukoharjo Regency, while the lowest LE achievement, with a value of 69.54 years, is in Brebes Regency.

Multiple Linear Regression Model LE in Central Java

In the initial stage of building the regression model, to identify which independent variables are related to LE, the following are the results of Pearson correlation analysis:

Table 1. Pearson Correlation

Variable	Value	p-value
X_1	-0.4303829	0.009859
X_2	0.4564822	0.005845
X_3	0.3004232	0.07951
X_4	-0.2875033	0.094
X_5	0.4412123	0.007975
X_6	-0.6186169	7.466×10^{-5}
X_7	0.2891654	0.09203

In Table 1, it is indicated that there are seven selected independent variables, namely: X_1, X_2, X_3, X_4, X_5, X_6, and X_7.

Table 2. Partial Significance Test of Linear Regression

	Estimate	t-value	p-value
Intercept	8.924×10^{1}	10.137	1.06×10^{-10}
X_1	-9.689×10^{-6}	-1.169	0.25266
X_2	2.103×10^{-2}	0.971	0.34015
X_3	-1.131×10^{-2}	-0.250	0.80431
X_4	-4.800×10^{-1}	-3.111	0.00437
X_5	1.031×10^{-4}	1.191	0.24389
X_6	-2.870×10^{-1}	-3.611	0.00123
X_7	-1.520×10^{-1}	-1.698	0.10101

The independent factors influencing life expectancy (LE) need to be identified. Identification of independent variables can be done by examining the reliability interval of parameters and the p-value from the t-test. If the parameter estimates include zero, the variable is not significantly influential. In this case, the influence of independent variables is examined using the p-value. In Table 2, it is shown that at a significant level (α) of 0.01 (10%), two variables independently significantly influence LE in Regencies and Cities in Central Java. The significant variables consist of the open unemployment rate per 1000 population (X_4) and the percentage of the population living in poverty (X_6). It can be observed that all p-values for the variables are less than α (0.1).

Formation of Spatial Weight Matrix

The Queen Contiguity weight matrix, also known as the WQ matrix (Queen Weight matrix), is utilized in spatial analysis to model the spatial relationships between neighboring areas or units. This matrix is employed to map the presence or intensity of neighbor relationships in spatial data. In the Queen Contiguity weight matrix, each area or unit is assigned a weight based on its neighbors. If two areas or units are adjacent, then the weight is 1, indicating the presence of a neighborly relationship. If two areas or units are not adjacent, then the weight is 0, indicating no neighborly relationship. This matrix defines Wij = 1 for entities that share a common side or vertex with the area of interest and Wij = 0 for other areas. An illustration of the spatial weight matrix is as follows:

$$W = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

The standardization of matrix results used in the model is:

$W_{ij}^{*} = \frac{W_{ij}}{\sum_{i=1}^{n} w_{ij}}$, Where $W_{ij(std)}$ is a weight matrix element that has been standardized

Testing Spatial Data Aspects

Spatial dependency test on residuals from linear regression using Moran's I. Moran's I is used to test the correlation between residuals and the regression line. The generated Moran's I value is 4.3773 (p-value = 0.000). The regression residual line indicates positive spatial autocorrelation at a significant level of 0.01. The Breusch-Pagan (BP) method is employed to test for heterogeneity. The resulting Breusch-Pagan (BP) value is 0.94643 (p-value = 0.3306), indicating no spatial heterogeneity or detected spatial homogeneity variance at the 0.01 level. The spatial regression model is performed after fulfilling the two spatial effect tests. Subsequently, Lagrange Multiplier (LM) testing is conducted to determine the appropriate spatial regression model. The LM test results are presented in Table 3 as follows:

Table 3. *Lagrange Multiplier* Test

	Value	p-value
Lagrange Multiplier lag (SAR)	5.2595	0,02183
Lagrange Multiplier error (SEM)	1.4317	0,23148
Lagrange Multiplier SARMA	6.5985	0,03691

The LM test results indicate that the SAR model is the most appropriate for testing the spatial dependence of LE in Central Java, based on the lowest p-value of 0.02183 among the three models. Therefore, LE in Central Java can be modeled using the Spatial Autoregressive (SAR) regression model.

***Spatial Autoregressive* (SAR)**

Table 4. Results of SAR Test

	Estiamate	z-value	p-value
Intercept	8.5080×10^{1}	11.6254	$<2.2 \times 10^{-16}$
X_1	-6.5694×10^{-6}	-0.9626	0.33573
X_2	4.6743×10^{-2}	2.2979	0.02157
X_3	1.2138×10^{-3}	0.0329	0.97378
X_4	-4.9205×10^{-1}	-3.9370	8.252×10^{-5}
X_5	1.0451×10^{-4}	1.4922	0.13564
X_6	-2.8512×10^{-1}	-4.4328	9.303×10^{-6}
X_7	-1.5232×10^{-1}	-2.1020	0.03555

Table 4 indicates that at a significance level of 0.01, there are 2 independent variables that significantly influence Life Expectancy (LE) in Central Java. The variables with significant values are the open unemployment rate (X_4) and the percentage of the population living in poverty (X_6).

Table 5. Parameter Estimation Results in SAR Model

	Estimate	Standar Error	z-value	p-value
Intercept	80.847296	0.998248	80.9892	$<2.2 \times 10^{-16}$
X_4	-0.334508	0.102310	-3.2696	0.001077
X_6	-0.356391	0.061661	-5.7799	7.475×10^{-9}

Based on table 5, the obtained SAR model equation for LE data in Central Java is as follows:

$$\hat{y}_i = 80.847296 - 0.334508X_4 - 0.356391X_6$$

to $$u_i = 0.55166 \sum_{j=1,i\neq j}^{n} w_{ij}u_j + \varepsilon_i$$

The results of the spatial autocorrelation coefficient (λ) being statistically significant spatially indicate the presence of spatial interdependence in the spatial residuals among one district/city with its neighboring district/city, amounting to 0.55166 multiplied by the average residual in its neighboring regions.

CONCLUSION

In this study, it can be concluded that based on the best spatial regression analysis, specifically the Spatial Autoregressive (SAR) model with Lagrange Multiplier (LM) testing, it was found that the Open Unemployment Rate (X_4) and Percentage of Poor Population (X_6) re factors influencing Life Expectancy (LE) in Central Java. These factors have a significant real impact at a level of 0.1 (10%). The Open Unemployment Rate (X_4) and Percentage of Poor Population (X_6) negatively affect Life Expectancy (LE) in Central Java.

REFERENCES

[1] Santika, N. Hanum, Safuridar, and Asnidar, "Pengaruh Jumlah Penduduk, Angka Harapan Hidup dan Rata-Rata Lama Sekolah terhadap Indeks Pembangunan Manusia di Kabupaten Aceh Tamiang," *Jurnal Ekonomi dan Manajemen*, vol. 2, no. 4, 2022.

[2] BPS, "Badan Pusat Statistik Kabupaten Semarang," Jun. 2023. Accessed: Jun.08, 2023. [Online]. https://semarangkab.bps.go.id/indicator/40/161/1/angka-harapan-hidup-ahh-menurutkabupaten-kota-dan-jenis-kelamin-di-jawa-tengah.html

[3] Z. Niaz Mahmud and K. Asif, "A Spatial Regression Modeling Framework for Examining Relationships Between the Built Environment and Pedestrian Crash Occurrences at Macroscopic Level: A Study in A Developing Country Context," *Geography and Sustainability*, vol. 3, no. 4, pp. 312–324, Dec. 2022, doi: 10.1016/j.geosus.2022.09.005.

[4] J. Olmo and M. Sanso-Navarro, "A Nonparametric Spatial Regression Model Using Partitioning Estimators," *Econom Stat*, Feb. 2023, doi: 10.1016/j.ecosta.2023.02.003.

[5] R. Evi, S. Nany, and M. Medina Suha, "Identifikasi Faktor-Faktor yang Memengaruhi Angka Harapan Hidup di Sumatera Tahun 2018 Menggunakan Analisis Regresi Spasial Pendekatan Area," 2020.

[6] F. Fatkhurohman, "Model Regresi Spasial Terbaik Indeks Pembangunan Manusia Provinsi Jawa Tengah," 2016.

[7] K. Suryowati, R. D. Bekti, and A. Faradila, "A Comparison of Weights Matrices on Computation of Dengue Spatial Autocorrelation," in *IOP Conference Series: Materials Science and Engineering*, Institute of Physics Publishing, Apr. 2018. doi: 10.1088/1757-899X/335/1/012052.

[8] W. S. Tarigan, "Analisis Regresi Spasial pada Indeks Pembangunan Manusia di Provinsi Sumatera Utara Tahun 2020 (Spatial Regression Analysis on the HDI in North Sumatera Province in 2020)," 2020.

[9] L. Ni Made Lasti, S. I Wayan, and S. I Komang Gede, "Pemodelan Jumlah Tindak Kriminalitas Di Provinsi Jawa Timur Dengan Analisis Regresi Spatial Autoregressive And Moving Average," *E-Jurnal Matematika*, vol. 7, no. 4, p. 346, Dec. 2018, doi: 10.24843/mtk.2018.v07.i04.p224.

[10] Firmansyah, Rangga Hadi. (2022). 5 Provinsi dengan Angka Harapan Hidup Tertinggi di Indonesia. *https://goodstats.id/article/5-provinsi-dengan-angka-harapan-hidup%20tertinggi-di-indonesia-xz4cQ*.

[11] Halicioglu, F. (2011). Munich Personal RePEc Archive Modelling life expectancy in Turkey. *Economic Modelling*, *28*(5), 2075–2082. https://doi.org/10.1016/j.econmod.2011.05.002.

[12] World Population Data Sheet. (2020). *Demographic Trends May Make Us Vulnerable to Pandemics Data Table*. 22. https://www.prb.org/wpcontent/uploads/2020/07/letter-booklet-2020-world-population.pdf

[13] Dindas Kesehatan Jayapura. Diakses pada tanggal 7 Mei 20023. *https://dinkes.jayapurakab.go.id/2933-2/#:~:text=H*.

[14] A. Yasir *et al.*, "Model Regresi Spasial untuk Analisis Presentase Penduduk Miskin di Provinsi Nanggroe Aceh Darussalam," *Jurnal Statistika Industri dan Komputasi*, vol. 1, no. 1, pp. 53-61, 2016.

[15] R. Faizatun Nisa and A. Rachman Hakim, "Pemodelan Mixed Geographically Weighted Regression dengan Adaptive Bandwidth untuk Angka Harapan Hidup (Studi Kasus: Angak Harapan Hidup di Jawa Tengah)," vol. 11, no. 1, pp. 67-76, 2022, [Online]. Available: https://ejournal3.undip.ac.id/index.php/gaussian/

[16] Y. Wardani, "Estimasi Parameter *Spatial Error Model* yang Memuat Pecilan," 2019.

[17] Chotimah Husnul and I. Rinjani, "Pemodelan Spasial Konsumsi Pemerintah dalam Perekonomian Jawa Timur: Spatial Autoregressive and Moving Average," *Jurnal Ilmiah Komputasi dan Statistika*, vol. 2, pp. 2087-3657, 2022

[18] Alfiani, S., Arum, P. R., & Arum, R. (2022). Pemodelan Pertumbuhan Ekonomi di Jawa Barat Menggunakan Metode Geographically Weighted Panel Regression. In *Universitas Muhammadiyah Semarang Jl. Kedungmundu* (Vol. 15, Issue 2). www.unipasby.ac.id

[19] Huriyatullah Rona Nabila, N., Fitri, Y., Rismawati Arum, P., Studi Statistika, P., & Matematika dan Ilmu Pengetahuan Alam, F. (2023). *Analisis Faktor-Faktor Yang Mempengaruhi Indeks Pembangunan Manusia Berdasarkan Kabupaten/Kota Di Jawa Tengah* (Vol. 16, Issue 1).

[20] Oktaviana, E., Arum, P. R., & Al Haris, M. (n.d.). *Pemodelan Spatial Autoregressive Quantile Regression (SARQR) Menggunakan Pembobot Queen Contiguity Pada Kasus Stunting Balita di Indonesia Spatial Autoregressive Quantile Regression (SARQR) Modeling Using Queen Contiguity Weights in Toddler Stunting Cases in Indonesia*.

[21] *Demographic Trends May Make Us Vulnerable to Pandemics Data Table*. (n.d.).