Parameter: Journal of Statistics
Available at https://bestjournal.untad.ac.id/index.php/parameter
Vol. 5 No. 1 2025, 13 - 26
DOI: *https://doi.org/10.22487/27765660.2025.v5.i1.17397*
eISSN: 2776-5660

SPATIAL AUTOREGRESSIVE MODEL (SAR) AND SPATIAL ERROR MODEL (SEM) MODELING ON LIFE EXPECTANCY DATA IN SOUTH SULAWESI PROVINCE 2022

Ayu Pebriyanti[1], Hardianti Hafid[2*], Sudarmin[3]

[1,2,3]Universitas Negeri Makassar

e-mail: [1]ayupbrynt21@gmail.com, [2*]hardiantihf@unm.ac.id, [3]sudarmin@unm.ac.id

ABSTRACT

Spatial regression is a development of classical linear regression which takes into account the spatial or spatial effects of the data being analyzed. The Spatial Autoregressive Model (SAR) and Spatial Error Model (SEM) methods include spatial regression models show that spatial effects on response variables and predictor variables. This research aims to model the factors that influence life expectancy in South Sulawesi Province in 2022. The analysis method used in this research is the SAR and SEM methods. The results show that based on the Lagrange Multiplier test values, there are lag and error dependencies. Based on the research results, it was found that the SAR and SEM models each had Akaike's Information Criterion (AIC) values of 94,0069 and 90, 6410, so the best model for analyzing the influence life expectancy value was the SEM model because the smallest had Akaike's Information Criterion (AIC) value was obtained. The factors that have a significant influence on life expectancy are average years of schooling and gross regional domestic product which have a positive effect. Then, the percentage of poor population and per capita expenditure have a negative effect.

Keywords: Spatial Regression, Spatial Autoregressive Model (SAR), Spatial Error Model (SEM), Life Expectancy

Cite: Pebriyanti, A., Hafid, H., & Sudarmin. (2025). Spatial autoregressive model (SAR) and spatial error model (SEM) modeling on life expectancy data in South Sulawesi Province 2022. *Parameter: Journal of Statistics*, *5*(1), 13–16. https://doi.org/10.22487/27765660.2025.v5.i1.17397

*Received: May 2025; Reviewed: May 2025; Published: Jul 2025

INTRODUCTION

Regression is a statistical analysis technique used to understand the relationship between a dependent variable and one or more independent variables, and to predict the value of the dependent variable based on the independent variables (Mukrom et al., 2021). One type of regression analysis is spatial regression, which is an extension of classical linear regression that accounts for location effects its analysis. Spatial data contains information about location, allowing analysis to consider not only the data itself but also its geographic position (Sunandi et al., 2021).

Spatial regression is often used when there is a correlation between adjacent locations, known as spatial dependency. In this research, area data types were used because the analysis focuses on regions (districts/cities) and their spatial relationships. Area data effectively capture the geographic contiguity and spatial interactions between neighboring administrative regions, which is essential in identifying spatial dependencies in life expectancy. The Spatial Autoregressive Model (SAR) and the Spatial Error Model (SEM) are two types of spatial regression. SAR combines linear regression with spatial lag on the response variable, while SEM is used when there is a correlation between error values at different locations. Both of these models can be used for analyzing phenomena with spatial aspects, such as life expectancy (Arum et al., 2023). Life expectancy is the average age that a population can expect to reach within a certain period. It reflects the level of public health and is often used as an indicator of government success in improving population well-being. Life expectancy is closely related to the quality of healthcare services—better health conditions tend to lead to longer life expectancy. In Indonesia, life expectancy in 2022 reached 73.6 years, a slight increase from the previous year, suggesting improvements in health quality and socio-economic conditions (BPS, 2023). In this study, life expectancy is used as the dependent variable. The use of spatial econometric models such as Spatial Autoregressive (SAR) and Spatial Error Model (SEM) is motivated by the assumption that life expectancy across districts in South Sulawesi may exhibit spatial dependence. This assumption is reasonable because neighboring districts often share similar health infrastructure, socio-economic environments, and public service quality, all of which can influence the health outcomes of the population. Therefore, spatial interactions between regions should be considered in the analysis. Previous studies have also shown that life expectancy clusters geographically. For instance, Hakim et al. (2019) modeled life expectancy in Central Java using a Spatial Durbin Model (SDM) and found significant spatial effects. Likewise, Yasin et al. (2022) developed a modified SAR for life expectancy in Central Java, demonstrating improved model accuracy and highlighting the benefit of spatial econometrics in capturing regional dependencies.

In various provinces of Indonesia, life expectancy shows variation, and factors such as environment, education, and health can influence it. For example, in South Sulawesi, according to data from Central Statistics Agency, life expectancy increased from 73,5 years in 2020 to 73,6 years in 2022. An increase or decrease in life expectancy in a region reflects the impact of local factors on the health and well-being of the community. Previous studies on SAR and SEM, such as the research by (Rahmawati & Bimanto, 2021), which compared SAR and SEM in modeling the Human Development Index in East Java Province, showed that the Lagrange Multiplier test indicated lag and error dependencies. The study found that the SEM model had the highest R^2 and the lowest Akaike's Information Criterion (AIC) value, indicating that SEM is a better choice for modeling the Human Development Index in East Java Province compared to the SAR model OLS regression. The study by (Dewi et al., 2019) which compared multiple regression, Spatial Autoregressive Model (SAR), and Spatial Error Model (SEM) for malnutrition in Indonesia In 2017, showed a comparison between multiple regression and spatial regressions like SAR and SEM. Based on the comparison results, the best model was the Spatial Autoregressive Model (SAR), with a determination coefficient of 63,1379% and the smallest AIC value of 99,2843. The study by Wasono et al. (2018), which focused on the planning of School Operational Assistance programs in Central Java Province using Spatial Autoregressive (SAR) and Spatial Error Model (SEM), Showed that the SEM model was superior to the SAR model due to its lower Akaike's Information Criterion (AIC) value of 60,3440. This research provides a new perspective by modeling life expectancy in South Sulawesi Province using spatial regression techniques, an area of study that has rarely been explored in previous research. Unlike earlier studies that applied SAR and SEM to various indicators such as the Human Development Index or malnutrition rates, this study specifically examines spatial dependencies in life expectancy at the district/city level in South Sulawesi. This regional focus provides valuable insights for local policy development related to public health and regional inequality.

This research contributes to the existing literature by specifically applying spatial regression models to analyze life expectancy at the district and city level in South Sulawesi Province, which has rarely been examined in previous studies. While prior research has explored spatial dependencies in various socio-economic indicators such as the Human Development Index or malnutrition rates in other regions of Indonesia, studies focusing on spatial modeling of life expectancy in South Sulawesi remain limited. The novelty of this study lies in its regional focus, the use of up-to-date 2022 data, and the application of both SAR and SEM models to compare their suitability for modeling life expectancy. The findings are expected to provide new insights into the spatial dynamics of public health in South Sulawesi and support evidence-based policy formulation aimed at reducing regional disparities in health outcomes. Based on this background, the aim of this study is to model the factors affecting life expectancy in South Sulawesi Province in 2022 using the Spatial Autoregressive Model (SAR) and Spatial Error Model (SEM) methods.

MATERIALS AND METHODS

Regression Analysis

Regression analysis is a method used to examine the relationship between a response variable and one or more explanatory variables. Multiple regression equations involve a single response variable (Y) affected by several predictor variables. The relationship between these variables can be formulated in a model:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_n X_{ni} + \varepsilon_i \tag{1}$$

where, Y is the response variable (the predicted value), X is the predictor variable, β_0 is the constant, β_n is the regression coefficient of the n predictor variable and ε is the error. In matrix form, the equation can be written as:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{2}$$

where, $\boldsymbol{y}$ is a vector of size $(n \times 1)$, $\boldsymbol{X}$ is a matrix of size $n \times (p \times 1)$, $\boldsymbol{\beta}$ is a vector of size $(p \times 1) \times 1$ and $\boldsymbol{\varepsilon}$ is a vector of size $(n \times 1)$, after determining the regression model, a classical assumption test is conducted. The assumptions underlying the regression model are:

a. Normality Test

The normality test is used to determine whether residuals in the regression equation are normally distributed. The normality test can be conducted using the Kolmogorov-Smirnov test (Massey, 1951)

$$D = \max |\, F_0\,(x_i\,) - S_n(x_i\,)|,\ i = 1,2, \ldots\ldots, n \tag{3}$$

where $F_0(x_i\,)$ refres to the cumulative distribution function of the theoretical distribution under H_0, $S_n(x_i\,)$referes to the cumulative frequency distribution of n observations, and $H_0$the residuals that follow a normal distribution. The decision criteria for the Kolmogorov-Smirnov test are that if the $D < D_{table}$ value or if the $p - value$ is greater than the significance level α, then the assumption of normality is satisfied.

b. Multicollinearity Test

Multicollinearity occurs when there is a high correlation among independent variables. The Variance Inflation Factor (VIF) is one method to measure the extent of multicollinearity (Theil, 1971)

$$VIF_j = \frac{1}{1 - R_j^2} \tag{4}$$

where R_j^2 refers to the coefficient of determination obtained from the regression of X_j on the other independent variables. If the $VIF_i < 10$ value, it is accepted, meaning there is no multicollinearity.

c. Heteroscedasticity Test

The heteroscedasticity test is conducted to examine whether there is an inequality in the variance is not constant, it is referred to as heteroscedasticity. One method detect heteroscedasticity is by using the Breusch-Pagan test. The testing criterion in the Breusch-Pagan test (Breusch & Pagan, 1979) is to reject H_0, if the $BP > \chi^2$ value or $p - value > \alpha$.

$$BP = \frac{1}{2}\, \boldsymbol{f}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}\boldsymbol{f} \tag{5}$$

where elements of factor $\boldsymbol{f}$ are $f_i = \frac{e_i^2}{\sigma^2} - 1$, e_i is the i_{th} residual from the multiple linear regression results.

d. Autocorrelation Test

The assumption of independence or the residual autocorrelation test is conducted to determine whether there is any correlation in the residual data. The test used is the Durbin-Watson test. The Durbin-Watson (Durbin & Watson, 1951):

$$d = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=2}^{n} e_t^2}$$

Where d denotes the Durbin-Watson value, e_t represents the residual at time t, and e_{t-1} refers to the residual at the previous time period (one lag). Test result will reject H_0 if the $d < d_u$ or $d > (4 - d_u)$. In addition to this test, the $p - value$ can be also be considered. If the $p - value > \alpha$, it means H_0 is accepted, indicating that there is no autocorrelation in the regression model.

Weight Matrix

A spatial weight matrix is used to identify the proximity or relationship between spatial data. This matrix is used to calculate the autocorrelation coefficient in the observed area. The matrix consists of entries representing the weight values assigned for comparing the observed areas. In spatial analysis, the spatial weight matrix is crucial because it represents the relationships between observation areas of size $n \times n$ and is usually denoted by $\boldsymbol{W}$. The weighting matrix used in this study is standardized- Queen Contiguity, which is a contiguity concept that assigns a value of 1 to regions that share either a border or a corner with the observed region, and a value of 0 to all other regions. An illustration of the spatial weights matrix using Queen Contiguity is as follows (LeSage, 1999)

$$\boldsymbol{W} = \begin{bmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{n1} & \cdots & w_{nn} \end{bmatrix} \tag{6}$$

Since the sum of the rows must equal 1 ($\sum_{i=1}^{n} w_{ij}$ = 1 for i = 1 to n) the matrix W must be standardized. The standardized matrix is the one that will be used in the model, where

$$\boldsymbol{W}^* = \frac{w_{ij}}{\sum_{i=1}^{n} w_{ij}} \tag{7}$$

Moran's I Index

Moran's I Index measures whether variables x (x_i and x_j), where $i \neq j, i = 1, 2, \dots, n; j = 1, 2, \dots, n$ with a dataset of size n, the formula for Moran's Index is (Lee & Wong, 2001):

$$I = \frac{N\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{(\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij})\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{8}$$

where, N refers to the number of observations, $\bar{x}$ represents the average of the observed variable, x_i refers to the observed value at location i, while x_j refers to the observed value at location j. and w_{ij} is spatial weight indicates the weighting factor assigned based on the spatial relationship between locations. The value of this index ranges from -1 to 1. A value of $-1 \leq I < 0$ indicates negative spatial autocorrelation, while a value of $0 < I \leq 1$ indicates positive spatial autocorrelation.
The expected value of Moran's Index is:

$$E(I) = I_0 = \frac{1}{n-1} \tag{9}$$

The presence or absence of autocorrelation in the data is assessed by comparing the value of Moran's Index (I) with the expected value of Moran's Index (I_0). The hypotheses are as follows:

$H_0: I = I_0$ (no spatial correlation between regions)

$H_1: I \neq I_0$ (spatial correlation between regions exists)

Statistic test formula:

$$\mathrm{Z} = \frac{\mathrm{I-E(I)}}{\mathrm{S_{error\,(I)}}} \tag{10}$$

where I is the value of Moran's I index, $E(I)$ is the expected value of I, and$S_{\text{error (I)}}$ is the standard error of I. Moran's I is considered statistically significant if the Z-test statistic is greater than the critical value $Z_{0.05}$.

Spatial autocorrelation is considered positive if $I > I_0$ and the data pattern is cluterd. If $I < I_0$, it indicates negative autocorrelation and the data pattern is dispersed. If $I = I_0$, it means there is no autocorrelation in the data.

Visually, the Moran's I scatter plot is used to evaluate the spatial relationship between observation values from neighboring locations. This diagram can be divided into four quadrants, which are:

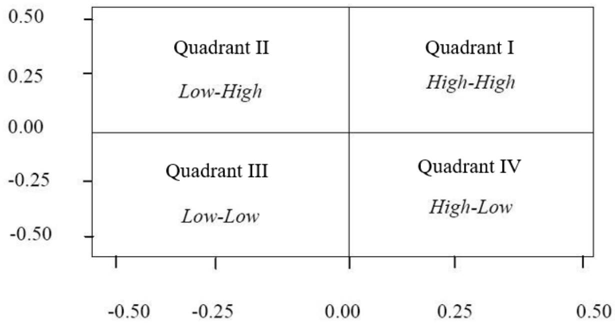

Figure 1. Moran's I Scatter Plot Illustration (Yuriantari et al., 2017)

Quadrant I (High-High) indicates locations with high observation values surrounded by other locations with high observation values. This quadrant represents a positive spatial autocorrelation were high values cluster together. Quadrant II (Low-High) indicates locations with low observation values surrounded by locations with high observation values. This quadrant represents a negative spatial autocorrelation, where low values are adjacent to high values. Quadrant III (Low-Low) indicates locations with low observation values surrounded by other locations with low observation values. This quadrant represents a positive spatial autocorrelation where low values cluster together. Quadrant IV (High-Low) indicates locations with high observation values surrounded by location with low observation values. This quadrant represents a negative spatial autocorrelation, where high values are adjacent to low values (Yuriantari et al., 2017)

Spatial Dependence Test

The spatial dependence effect, based on Tobler's First Law, states that closer objects have a greater influence in each other. To address the spatial dependence effect, area-based approaches and the Lagrange Multiplier (LM) test are used for detection (Yasin et al., 2020). The hypothesis testing for the Lagrange Multiplier (LM) is as follows:

a. Spatial Autoregressive Model (SAR)

The hypotheses for the SAR model are as follows:

$H_0: \rho = 0$ (no spatial lag dependence)

$H_1: \rho \neq 0$ (spatial lag dependence exixts)

The test statistic used is:

$$LM_\rho = \frac{\left[\dfrac{\boldsymbol{\varepsilon}'\boldsymbol{W}\boldsymbol{y}}{\left(\dfrac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{n}\right)}\right]^2}{D} \tag{11}$$

where,

$$D = \left[\frac{(\boldsymbol{WX\beta})'(I - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{WX\beta})}{\sigma^2}\right] + tr(\boldsymbol{W}'\boldsymbol{W} + \boldsymbol{WW})$$

where, $\boldsymbol{\varepsilon}$ is the error vector of the regression model, n is the number of observations, $\boldsymbol{W}$ is the the spatial weight matrix of size $n \times n$, $\boldsymbol{X}$ is the matrix of independent variables of size $n \times (p+1)$, $\boldsymbol{y}$ is the response variable matrix of size $n \times 1$, $\boldsymbol{\beta}$ is the vector of regression coefficient of size $(p+1) \times 1$, I is the identity matrix and σ^2 is the estimated variance of the regression model errors.

Reject H_0 if $LM_\rho > \chi^2{}_{(q)}$ or if the $p-value < \alpha$, which indicates the presence of spatial lag dependence.

b. Spatial Error Model (SEM)

The hypotheses for the SAR model are as follow:

$H_0: \lambda = 0$ (no spatial error dependence)

$H_1: \lambda \neq 0$ (spatial error dependence exists)

The test statistic used is:

$$LM_\lambda = \frac{\left[\frac{\boldsymbol{\varepsilon}' W \boldsymbol{\varepsilon}}{\left(\frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{n}\right)}\right]^2}{tr(\boldsymbol{W}'\boldsymbol{W} + \boldsymbol{W}\boldsymbol{W})} \tag{12}$$

where, $\boldsymbol{\varepsilon}$ is the error vector of the regression model, n is the number of observations, $\boldsymbol{W}$ is the spatial weight matrix of size $(n \times n)$.

Reject H_0 if $LM_\lambda > \chi^2{}_{(q)}$ or $p-value < \alpha$, which indicates the presence of spatial error dependence.

Spatial Regression

Spatial regression is an analytical method that evaluates the relationship between one variable and several other variables while considering the spatial effects at various locations that are the focus of observation. According to (J. P. LeSage, 1999) , the general model for spatial regression can be expressed as follows:

$$\boldsymbol{y} = \rho \boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}, \qquad u = \lambda W u + \varepsilon, \qquad \varepsilon \sim N(0, \sigma_\varepsilon^2 I_n) \tag{13}$$

where, $\boldsymbol{y}$ is the response variable vector of size $n \times 1$, $\boldsymbol{X}$ is the predictor variable matrix of size $n \times (p \times 1)$, β is the regression parameter coefficient vector of size $(p+1) \times 1$, $\boldsymbol{W}$ is the spatial weight matrix of size $n \times n$, u is the error vector, assumed to contain autocorrelation, of size $n \times 1$, ρ is the spatial lag autoregressive coefficient, λ is the spatial error autoregressive coefficient and I is the identity matrix of size $n \times n$.

Spatial Autoregressive Model (SAR)

The Spatial Autoregressive (SAR) model is formed when $\rho \neq 0$ and $\lambda = 0$, implying that the model assumes an autoregressive process in the responsen variable. The following is the equation for the Spatial Autoregressive (SAR) model is expressed as follows LeSage & Pace (2009):

$$\boldsymbol{y} = \rho \boldsymbol{W}_{\boldsymbol{y}} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad \varepsilon \sim N(0, \sigma_\varepsilon^2 I) \tag{14}$$

where, $\boldsymbol{y}$ is the response variable matrix of size $(n \times 1)$, $\boldsymbol{X}$ is the matrix of explanatory variables of size $(n \times (p+1))$, $\boldsymbol{W}$ is the spatial weught matrix of size $(n \times n)$, ρ is the spatial lag autoregressive coefficient, $\boldsymbol{\beta}$ is the vector of regression parameter coefficients of size $(p+1) \times 1$, ε is the error vector of size $(n \times 1)$, normally distributed with mean zero and variance $\sigma^2 I$.

Spatial Error Model (SEM)

The Spatial Error Model (SEM) is formed when $\lambda \neq 0$ and $\rho = 0$, assuming that the autoregressive process is only in the model's errors. The SEM is expressed as follows:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \lambda \boldsymbol{W}\boldsymbol{u} + \boldsymbol{\varepsilon}, \qquad \varepsilon \sim N(0, \sigma_\varepsilon^2 I) \tag{15}$$

where, y is the response variable matrix of size $(n \times 1)$, X is the matrix of explanatory variables of size $(n \times (p+1))$, $\boldsymbol{W}$ is the sptial weight matrix of size $(n \times n)$, $\boldsymbol{\beta}$ is the vector of regression parameter coefficients of size $(p+1) \times 1$, λ is the autoregressive lag coefficient on the error term, with $|\lambda| < 1$, $\boldsymbol{u}$ is the error vector of size $(n \times 1)$, normally distribused with mean zero and variance $\sigma^2 I$.

Selection of the Best Model

Akaike's Information Criterion (AIC) is a criterion for selecting the best model introduced by Akaike in 1973, which takes into account the number of parameters in the model. The AIC criterion can be formulated as follows:

$$\boldsymbol{AIC = n \times \ln\left(\frac{SSE}{n}\right) + 2f + n + n \times \ln(2\pi)} \tag{16}$$

where, ln is the natural logarithm, SSE is the Sum of Squared Errors, n is the number of observations, f is the number of parameters in the model, and π is approximately 3,14.
A smaller AIC value indicates a better model, as it suggests a better balance between the goodness of fit and model complexity (Aswi & Sukarna, 2006).

Life Expectancy

According to BPS (2023), life expectancy is the average number of additional years a person is expected to live once they have reached age x in a given year. Life expectancy serves as a tool to evaluate government performance in improving the overall well-being of the population and enhancing health outcomes in specific cases. Factors that are believed to influence life expectancy include educational factors, health factors, and socio-economic factors.

Data and Data Sources

The data used in this study are secondary data obtained from the Central Statistics Agency and the Directorate General of Village Governance, Ministry of Home Affairs, regarding life expectancy, average years of schooling, school participation rates, Gross Regional Domestic Product (GRDP), percentage of the poor participation, per capita expenditure, and number of posyandu in the regencies/cities of South Sulawesi Province in 2022. The response variable and explanatory variables used in this study are:

Table 1. The Variables used

Code	Variable	Description
Y	Life Expectancy	The variable life expectancy is defined as the average number of years that an individual or member of a population in a particular region or country is expected to live.
X_1	Average Years of Schooling	Average years of schooling is a variable that measures the number of years of schooling attained by the target age population within a specific region.
X_2	School Participation Rate	The school participation rate is a variable that measures the proportion or percentage of the school-aged population that is currently attending or enrolled in formal educational institutions during a specific period of time.
X_3	Gross Regional Domestic Product	Gross regional domestic product is the total value of goods and services produced within a specific region over a certain period of time.
X_4	Percentage of the Poor Population	The poverty rate is a variable that measures the percentage of the total population in a region or country living below the poverty line.
X_5	Per Capita Expenditure	The per capita expenditure variable measures the average amount of spending by an individual.
X_6	Number of Posyandu	The variable that measures the number of public health centers.

Data Analysis Techniques

The data analysis techniques that will be used in this research are as follows:

1. Inputting the data into R Studio software.
2. Conducting classical regression tests using R Studio.
3. Determining the spatial weight matrix to show the neighborhood relationships between locations.
4. Performing Moran's Index dependency test. If there is a dependency, proceed with the Lagrange Multiplier (LM) test; if not, the process is complete.
5. Conducting the Lagrange Multiplier (LM) test.
6. If the LM test indicates spatial lag (LM_{lag}), estimate and test the significance of the parameters in the Spatial Autoregressive Model (SAR); if, not, the process is complete.

7. If the LM test indicates spatial error (LM_{error}), estimate and test the significance of the parameters in the Spatial Error Model (SEM); if not, the process is complete.
8. Selecting the best model.
9. Conclusion

RESULTS AND DISCUSSION

Descriptive Data Analysis

South Sulawesi is one of the provinces in Indonesia, located on the island of Sulawesi, and consists of 24 regencies/cities. The following are the results of the descriptive analysis:

Table 2. Descriptive Analysis of Observed Variables

Variable	Minimum	Maximum	Average
Life Expectancy	66,8100	73,7200	69,8900
Average Years of Schooling	6,7500	11,5500	8,3600
School Participation Rate	18,5200	55,1200	31,0900
Gross Regional Domestic Product	1,9900	15,4500	5,4600
Percentage of the Poor Population	4,5800	13,9200	9,3200
Per Capita Expenditure	7,584	17,406	11,272
Number of Posyandu	25	2135	249,8800

Based on Table 2, it is known that the life expectancy variable has a minimum value between 66-67 years, a maximum value between 73-74 years, and an average value between 69-70 years. For the average years of schooling variable, the minimum value is between 6-7 years, the maximum value is between 11-12 years, and the average value is between 8-9 years. For the school participation rate variable, the minimum value is 18,52%, the maximum value is 55,12%, and the average value is 31,09%. For the gross regional domestic product variable, the minimum value is 1,99%, the minimum value is 15,45%, and the average value is 5,46%. For the poverty rate variable, the minimum value is 4,58%, the maximum value is 13,92%, and the average value is 9,32%. For the per capita expenditure variable, the minimum value is 7,584,000 IDR, the maximum value is 17,406,000 IDR, and the average value is 11,272,000 IDR. For the number of posyandu variable, the minimum value is 25 units, the maximum value is 2,135 units, and the average value is between 249-250 units.

Classical Regression Testing

a. Normality Test

The normality test can be conducted using the Kolmogorov-Smirnov test with the following hypotheses:

H_0: Residuals are normally distributed

H_1: Residuals are not normally distributed

The test results showed a value of $D = 0{,}0193 < D_{table} = 0{,}2420$ and the $p - value$ = 0,6455. Therefore, H_0 is accepted, indicating that the residuals are normally distributed.

b. Multicollinearity Test

The multicollinearity test is used to determine whether there is any correlation among the predictor variables with the following hypotheses:

H_0: There is no multicollinearity

H_1: There is multicollinearity

This can be assessed by examining the Variance Inflation Factor (VIF) values, as shown in Table 3.

Table 3. Multicollinearity Test

Variable	X_1	X_2	X_3	X_4	X_5	X_6
VIF	3,8355	3,0574	1,3124	1,8033	2,3346	1,1626

Based on Table 3, it is observed that the VIF values for each predictor variable less than 10, which means that H_0 is accepted. This indicates that there is no multicollinearity present in the observed data.

c. Heteroskedasticity Test

To determine the presence of heteroskedasticity or variance differences across regions, a heteroskedasticity test is conducted using the Breusch-Pagan test with the following hypotheses:

H_0= There is no heteroskedasticity

H_1= There is heteroskedasticity

The results of the heteroskedasticity test are presented in Table 5.

Table 5. Breusch-Pagan Test

Breusch-Pagan test		
BP	**Df**	$\boldsymbol{P - Value}$
3,3269	6	0,7668

Based on Table 5, the Breusch-Pagan test value is 3,3269, which is greater than the criitical value of 0.0999, and the $p - value$ is 0,77668, which is greater than the significance level α = 10%. Therefore, H_0 is accepted, indicating that there is no heteroskedasticity.

d. Autocorrelation Test

The detection of autocorrelation is conducted using the Durbin-Watson Test with the following hypotheses:

H_0: There is no autocorrelation

H_1: There is autocorrelation

The results of the autocorrelation test are presented in Table 4.

Table 4. Autocorrelation Test

Test	**DW Value**	$\boldsymbol{P - Value}$
Durbin-Watson	1,6778	0,1634

Based on Table 4, the test result shows a values of $d_{calculated} = 1{,}6778 < d_u = 2{,}0352$ and a $p - value = 0{,}1634$, which is greater than the significance level α = 10%. This leads to the acceptance of H_0. Therefore, it can be concluded that there is no autocorrelation in the model.

Spatial Weight Matrix

The proximity or relationship between observation locations is expressed in the spatial weight matrix. In this study, the spatial weight matrix used is the standardized Queen Contiguity matrix. Queen Contiguity is a combination of Rook Contiguity and Bishop Contiguity. It is a contiguity matrix where the observed areas are adjacent at both sides and corners. Neighboring areas are assigned a weight of 1, while non-neighboring areas are assigned a weight of 0. The following is the standardized Queen Contiguity matrix used in this study.

$$\begin{bmatrix} 0 & 0{,}2 & 0 & \dots & 0 \\ 0{,}14 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0{,}33 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0{,}25 & \dots & 0 \end{bmatrix}$$

Moran's I Index

The Moran's I test is used to assess the presence of spatial dependence in the data. It evaluates whether the values at one location are correlated with values at nearby locations, indicating spatial clustering or dispersion.

Table 6. Results of Moran's Index Test

Moran's I Index	$\boldsymbol{P - Value}$
0,2384	0,0260

Based on Table 6, the $p - value$ is 0,0260, which is smaller than the significance level (α) of 10%. This indicates the presence of spatial dependence in the data. Additionally, the Moran's I Index value is 0,2384, which falls within the range of $-1 < I \leq 1$. This value suggests a clustered distribution pattern, where nearby areas have similar or identical values compared to more distant areas.

Moran Scatterplot

The Moran Scatterplot is used to detect or identify spatial dependence based on spatial quadrant types.

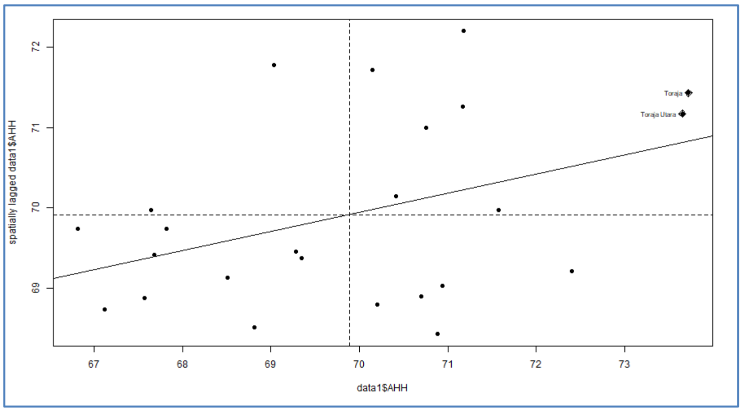

Figure 2. Moran Scatterplot for Life Expectancy in South Sulawesi in 2022

Based on Figure 2, it is found. The regencies/cities of Enrekang, Luwu, Palopo, Parepare, Pinrang, Sidrap, North Toraja, and Toraja are located in Quadrant I (High-High), which means these areas have a high life expectancy and are also surrounded by areas with high life expectancy. Luwu Utara and Takalar Regencies are in Quadrant II (Low-High), indicating that these areas have a low life expectancy but are surrounded by areas with high life expectancy. The regencies/cities of Barru, Bone, Bulukumba, Jeneponto, Maros, Pangkep, Selayar, Sinjai, and Wajo are in Quadrant III (Low-Low), which means these areas have a low life expectancy and are also surrounded by areas with low life expectancy. The regencies/cities of Gowa, Luwu Utara, Makassar, Soppeng, and Bantaeng are in Quadrant IV (High-Low), indicating that these areas have a high life expectancy but are surrounded by areas with low life expectancy.

Lagrange Multiplier (LM) Test

The Lagrange Multiplier (LM) test is used to select the appropriate spatial regression model.

Table 7. Result of the Lagrange Multiplier (LM)

	Statistic	**Parameters**	$P - Value$
LMerr	3,0180	1	0,0823*
LMlag	3,3129	1	0,0687*
SARMA	3,4101	2	0,1817*

Explanation: * = Significant at 0,1 level

Based on Table 7, the $p - value$ faor LMerr and LMlag are 0,0823 and 0,0687, respectively, which are smaller than the 10% significance level (α). This indicates that both the Spatial Autoregressive Model (SAR) and the Spatial Error Model (SEM) are appropriate spatial regression models to use.

Spatial Autoregressive Model (SAR)

Modeling the factors influencing life using the SAR model.

Table 8. Estimation of Parameters in the SAR Model

	Estimation	***Std. Error***	$z - value$	$p - value$
Constant	38,8894	13,7097	2,8366	0,0045
X_1	1,2422	0,3946	3,1477	0,0016
X_2	0,0056	0,0541	0,1037	0,9173
X_3	0,2619	0,1104	2,3717	0,0177
X_4	-0,1596	0,1152	-1,3850	0,1660
X_5	-0,0004	0,0001	-2,7916	0,0052
X_6	-0,0009	0,0006	-1,4543	0,1458

	Estimation	*Std. Error*	$z-value$	$p-value$
ρ	0,3764	0,1967	1,1931	0,0729

Based on Table 8, it is found that the variables X_1, X_3, and X_5 have p-values smaller than the 10% significance level (α). This indicates that average years of schooling (X_1), Gross Regional Domestic Product (X_3), and per capita expenditure (X_5) have a significant effect on life expectancy (Y) in South Sulawesi Province. On the other hand, the variables X_2, X_4, and X_6 have p-values greater than the 10% significance level (α). This suggests that school enrollment rate (X_2), poverty rate (X_4), and number of health posts (X_6) do not have a significant impact on life expectancy (Y) in South Sulawesi Province.

Thus, the resulting SAR model is as follows:

$$\hat{Y} = 0{,}3764Wy + 38{,}8894 + 1{,}2422X_1 + 0{,}2619X_3 - 0{,}0004X_5 + \varepsilon$$

The developed model can explain the influence of average years of schooling ($X1$), Gross Regional Domestic Product ($X3$), and Expenditure Per Capita ($X5$) on life expectancy (Y), as follows

a. The value of ρ = 0,3764 indicates the spatial influence of a region surrounded by neighboring regions. This means that the life expectancy (Y) of each neighboring region may increase by 0,3764.
b. The positive constant coefficient indicates that if other variables are considered constant, the life expectancy (Y) in South Sulawesi tends to increase by 38,8894.
c. The average years of schooling variable ($X1$) has a regression coefficient of 1,2422. The positive sign indicates a positive relationship between average years of schooling ($X1$) and life expectancy (Y). This means that if average years of schooling ($X1$) increases by one unit, while the other factors—Gross Regional Domestic Product ($X3$), expenditure per capita ($X5$), and the spatial weight matrix—are held constant, life expectancy (Y) will increase by 1,2422.
d. The Gross Regional Domestic Product variable ($X3$) has a regression coefficient of 0,2619. The positive sign indicates a positive relationship between Gross Regional Domestic Product ($X3$) and life expectancy (Y). This means that if Gross Regional Domestic Product ($X3$) increases by one unit, while the other factors—average years of schooling ($X1$), expenditure per capita ($X5$), and the spatial weight matrix—are held constant, life expectancy (Y) will increase by 0,2619.
e. The expenditure per capita variable ($X5$) has a regression coefficient of -0,0004. The negative sign indicates a negative relationship between expenditure per capita ($X5$) and life expectancy (Y). This means that if Expenditure Per Capita ($X5$) decreases by one unit, while the other factors—average years of schooling ($X1$), Gross Regional Domestic Product ($X3$), and the spatial weight matrix—are held constant, life expectancy (Y) will increase by 0,0004.

Spatial Error Model (SEM)

Modeling the factors influencing life expectancy using the SEM model.

Table 9. Estimation of Parameters in the SEM Model

	Estimation	*Std. Error*	$z-value$	$p-value$
Constant	67,4204	2,7184	24,8007	$< 2{,}2 \times 10^{-16}$
X_1	0,8944	0,3635	2,4604	0,0138
X_2	0,0470	0,0554	0,8482	0,3963
X_3	0,3117	0,0859	3,6290	0,0002
X_4	-0,2035	0,1026	-1,9833	0,0473
X_5	-0,0005	0,0001	-3,6801	0,0002
X_6	-0,0008	0,0005	-1,4396	0,1499
λ	0,6470	0,1542	4,1938	0,0103

Based on Table 9, the results show that the variables X_1, X_3, X_4 and X_5 have p-values smaller than the 10% significance level (α). This indicates that average years of schooling (X_1), Gross Regional Domestic Product (X_3), poverty rate (X_4), and per capita expenditure (X_5) have a significant effect on life expectancy (Y) in South Sulawesi Province. On the other hand, the variables X_2 and X_6 have p-values greater than the 10% significance level (α), suggesting that School Enrollment Rate (X_2) and Number of Health Posts (X_6) do not have a significant impact on life expectancy (Y) in South Sulawesi Province.

Thus, the resulting SEM is as follows:

$$\hat{Y} = 0{,}6470Wu + 67{,}4204 + 0{,}8944X_1 + 0{,}3117X_3 - 0{,}0235X_4 - 0{,}0005X_5 + \varepsilon$$

The resulting model explains the influence of average years of schooling ($X1$), Gross Regional Domestic Product ($X3$), percentage of poor population ($X4$), and per capita expenditure ($X5$) on life expectancy (Y), as follows:

a. The value of λ = 0,6470 indicates the influence of one region being surrounded by other regions. This means that the life expectancy (Y) in each neighboring region may increase by 0,6470.
b. The positive coefficient of the constant indicates that if all other variables are considered constant, the life expectancy (Y) in South Sulawesi tends to increase by 67,4204.
c. The variable average years of schooling ($X1$) has a regression coefficient of 0,8944. The positive sign indicates a positive relationship between average years of schooling ($X1$) and life expectancy (Y). This means that if average years of schooling ($X1$) increases by one unit while other factors—Gross Regional Domestic Product ($X3$), percentage of poor population ($X4$), per capita expenditure ($X5$), and the spatial weight matrix—are held constant, then life expectancy (Y) will increase by 0,8944.
d. The variable Gross Regional Domestic Product ($X3$) has a regression coefficient of 0,3117. The positive sign indicates a positive relationship between Gross Regional Domestic Product ($X3$) and life expectancy (Y). This means that if Gross Regional Domestic Product ($X3$) increases by one unit while other factors—average years of schooling ($X1$), percentage of poor population ($X4$), per capita expenditure ($X5$), and the spatial weight matrix—are held constant, then life expectancy (Y) will increase by 0,3117.
e. The variable percentage of poor population ($X4$) has a regression coefficient of -0,0235. The negative sign indicates a negative relationship between percentage of poor population ($X4$) and life expectancy (Y). This means that if the percentage of poor population ($X4$) decreases by one unit while other factors—average years of schooling ($X1$), Gross Regional Domestic Product ($X3$), per capita expenditure ($X5$), and the spatial weight matrix—are held constant, then life expectancy (Y) will increase by 0,0235.
f. The variable per capita expenditure ($X5$) has a regression coefficient of -0,0005. The negative sign indicates a negative relationship between per capita expenditure ($X5$) and life expectancy (Y). This means that if per capita expenditure ($X5$) decreases by one unit while other factors—Average Years of Schooling ($X1$), Gross Regional Domestic Product ($X3$), percentage of poor population ($X4$), and the spatial weight matrix—are held constant, then life expectancy (Y) will increase by 0,0005.

Selection of the Best Model

The researcher conducted a comparison between the results of the Spatial Autoregressive (SAR) model and the Spatial Error Model (SEM) to determine the best model. The best model was selected based on the Akaike's Information Criterion (AIC) value, the greater the confidence in the model.

Table 10. Determination of the Best Model

Model	AIC
SAR	94,0069
SEM	90,6417

Based on Table 10, the AIC value for the SEM model is smaller than that for the SAR model. Therefore, the best model to use for modeling life expectancy (Y) for South Sulawesi Province 2022 data is the SEM model. The selection of the Spatial Error Model (SEM) as the best model is based on its lower AIC value compared to the Spatial Autoregressive (SAR) model, indicating that the spatial dependence in the life expectancy data of South Sulawesi is more influenced by unobserved factors that are spatially distributed. The analysis shows that mean years of schooling and Gross Regional Domestic Product have a positive effect on life expectancy, while the percentage of poor population and per capita expenditure have a negative effect. These variables reflect socio-economic conditions and regional development disparities that are not always evenly distributed across regions, along with other potential factors such as healthcare service quality, public health behaviors, and infrastructure distribution, which are not fully captured in the model but are spatially correlated at the regional level. Therefore, the use of the SEM model is more appropriate because it can capture spatial autocorrelation arising from these latent variables, while the SAR model focuses more on direct interaction between neighboring regions. This finding also reflects the complexity of regional development dynamics that indirectly influence public health across administrative boundaries.

CONCLUSION

Based on the results and discussion of this study, the following conclusions can be drawn:

a. The analysis using the Spatial Autoregressive Model (SAR) for life expectancy (Y) yields the following model:

$$\hat{Y} = 0{,}3764Wy + 38{,}8894 + 1{,}2422X_1 + 0{,}2619X_3 - 0{,}0004X_5 + \varepsilon$$

b. The analysis using the Spatial Error Model (SEM) for life expectancy (Y) yields the following model:

$$\hat{Y} = 0{,}6470Wu + 67{,}4204 + 0{,}8944X_1 + 0{,}3117X_3 - 0{,}0235X_4 - 0{,}0005X_5 + \varepsilon$$

c. The analysis using the Spatial Autoregressive (SAR) and Spatial Error Model (SEM) for life expectancy (Y) yields Akaike's Information Criterion (AIC) values of 94,0069 for the SAR model and 90,6417 for the SEM model. This indicates that the SEM model is the best choice for modeling the data in this study, as it has the lowest AIC value. The significant factors affecting life expectancy (Y) are average years of schooling (X_1) and Gross Regional Domestic Product (X_3), both of which have a positive impact. Additionally, the poverty rate (X_4) and per capita expenditure (X_5) have a negative impact.

REFERENCES

Arum, P. R., Gautama, R. P., Fitriani, I., & Naza, F. (2023). *Identifying Factors that Influence Life Expectancy in Central Java Using Spatial Regression Models*. *16*(2), 605–613.

BPS. (2023). *Indikator Kesejahteraan Rakyat Indonesia 2023*. https://www.bps.go.id/id/publication/2023/11/06/7807339c2dfaed0ca8e0beaa/indikator-kesejahteraan-rakyat-2023.html

Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroskedasticity and random coefficient variation. *Econometrica*, *47*(5), 1287–1294. https://doi.org/https://doi.org/10.2307/1911963

Dewi, D. K., Azizah, A., & Widodo, E. (2019). Perbandingan Metode Regresi Berganda, Spatial Autoregressive dan Spatial Error Model Terhadap Gizi Buruk di Indonesia Tahun 2017. *Prosiding Konferensi Nasional Penelitian Matematika Dan Pembelajarannya*.

Durbin, J., & Watson, G. S. (1951). Testing for serial correlation in least squares regression, II. *Biometrika*, *38*((1–2)), 159–179. https://doi.org/https://doi.org/10.1093/biomet/38.1-2.159

Hakim, A. R., Yasin, H., & Rusgiyono, A. (2019). Modeling Life Expectancy In Central Java. *Media Statistika*, *12*(2), 152–163. https://doi.org/10.14710/medstat.12.2.152-163

Lee, J., & Wong, S. (2001). *Statistical Analysis with Arcview GIS*. John Wiley and Sons.

LeSage, J. P. (1999). *The Theory and Practice of Spatial Econometrics*. University of Toledo. https://doi.org/https://doi.org/10.2307/2553707

LeSage, J., & Pace, R. K. (2009). *Introduction to Spatial Econometrics*. New York: Chapman and Hall/CRC.

Massey, F. J. J. (1951). The Kolmogorov–Smirnov test for goodness of fit. *Journal of the American Statistical Association*, *46*(253), 68–78. https://doi.org/https://doi.org/10.1080/01621459.1951.10500769

Mukrom, M. H., Yasin, H., & Hakim, A. R. (2021). Pemodelan Angka Harapan Hidup Provinsi Jawa Tengah Menggunakan Robust Spatial Durbin Model. *Jurnal Gaussian*, *10*(1), 44–54. https://doi.org/10.14710/j.gauss.v10i1.30935

Rahmawati, D., & Bimanto, H. (2021). Perbandingan Spatial Autoregressive Model dan Spatial Error Model dalam Pemodelan Indeks Pembangunan Manusia di Provinsi Jawa Timur. *Jurnal Statistika Dan Aplikasinya*, *5*(1), 41–50.

Sunandi, E., Djuraidah, A., Sumertajaya, I. M., & Wigena, A. (2021). Dynamic Spatial Durbin Modeling (Case : Percentage of Poor Population in Dynamic Spatial Durbin Modeling (Case : Percentage of Poor Population in Indonesia 2010-2019). *Journal of Physics: Conference Series*, *1863*(1), 1–12. https://doi.org/10.1088/1742-6596/1863/1/012046

Theil, H. (1971). *Principles of econometrics*. New York: John Wiley & Sons.

Wasono, R., Karim, A., Darsyah, M. Y., & Suwardi. (2018). Perencanaan Program Bantuan Operasional Sekolah (BOS) di Provinsi Jawa Tengah Berbasiskan Model Spatial Autoregressive (SAR) dan Spatial Error Model (SEM). *Seminar Nasional Edusaintek*, *2012*, 1–4.

Yasin, H., Warsito, B., & Hakim, A. R. (2020). *Regresi Spasial (Aplikasi dengan R)*. WADE Group.

Yasin, H., Warsito, B., Hakim, A. R., & Azizah, R. N. (2022). Life Expectancy Modeling Using Modified Spatial. *Media Statistika*, *15*(1), 72–82. https://doi.org/10.14710/medstat.15.1.72-82

Yuriantari, N. P., Hayati, M. N., & Wahyuningsih, S. (2017). Analisis Autokorelasi Spasialtitik Panas Di Kalimantan Timur Menggunakan Indeks Moran dan Local Indicator Of Spatial Autocorrelation (LISA) Analysis Spatial Autocorrelation Hotspot in East Kalimantan Using Index Moran and Local Indicator of Spatial Autoco. *Jurnal EKSPONENSIAL*, *8*(1), 63–70.