Pinnacle Academic Press Proceedings Series

Vol. 1 2025

Article Open Access

Exploring the Influencing Factors of Life Expectancy in Various Regions of China based on Multiple Linear Regression

Zixuan Yang [1,*]

[1] The Experimental High School Attached to Beijing Normal University, 100032, Beijing, China
* Correspondence: Zixuan Yang, The Experimental High School Attached to Beijing Normal University, 100032, Beijing, China

Received: 10 March 2025
Revised: 17 March 2025
Accepted: 02 April 2025
Published: 07 April 2025

Abstract: Life expectancy is an important indicator for measuring the economic development level of a country or region. This study uses multiple linear regression to analyze the factors influencing life expectancy across different regions of China. The research finds that altitude, distance from the nearest coastline, hospital beds per 1000 people, per capita education expenditure, sulfur dioxide emissions, and the proportion of urban population all have varying degrees of impact on life expectancy in China's regions. Corresponding policy recommendations are also proposed.

Keywords: life expectancy; multiple linear regression; influencing factors

1. Introduction

1.1. Research Background

Life expectancy, also known as expected lifespan, refers to the average number of years a newborn (age 0 group) is expected to live [1]. Commonly abbreviated as life expectancy or average lifespan, it serves as a universal indicator to measure the economic development level, quality of life, and healthcare conditions of a country or region at a specific point in time. Life expectancy is also a critical metric for evaluating the health levels of residents across different regions and periods. Factors such as socioeconomic conditions and healthcare standards significantly influence human longevity, leading to substantial variations in life expectancy across societies and historical periods. Additionally, individual differences in physical constitution, genetic factors, and living conditions result in significant disparities in lifespan among individuals. Life expectancy, as a comprehensive indicator of public well-being, is one of the key metrics used by the World Health Organization (WHO) to measure health outcomes [2].

1.2. Research Significance

The *Healthy China 2030* blueprint released by the Chinese government in 2016 set a strategic goal of increasing the national average life expectancy to 79 years by 2030 [3]. Achieving this target relies on a comprehensive understanding of the patterns and mechanisms influencing changes in life expectancy. In the context of life expectancy research in China, exploring regional disparities and their differentiation mechanisms holds both theoretical and practical significance. On the one hand, China's vast territory encompasses regions with immense variation in natural and social conditions, potentially leading to significant regional differences in life expectancy. Furthermore, since the initiation of the reform and opening-up policy, China's social, economic, and ecological indices have evolved rapidly alongside social progress. The significance of these regional differences and the diversity of influencing factors provide valuable opportunities for an in-depth evaluation of the mechanisms driving life expectancy. On the other hand, narrowing internal gaps may become a key driver of life expectancy growth in China over the next few decades. Identifying lagging regions, time periods, and influencing factors is crucial for developing targeted policies to effectively improve overall life expectancy in China.

1.3. Research Objectives

This study aims to analyze the spatial and temporal differences in life expectancy across China by constructing a regression model that incorporates indicators such as geographic location, healthcare standards, education levels, economic development, ecological environment, and urbanization. Furthermore, the study investigates the influence of these indicators on life expectancy. The findings are expected to provide targeted recommendations for China's public welfare policies, accelerating improvements in public health, and narrowing internal disparities.

2. Problem Analysis

2.1. Indicator Selection

2.1.1. Response Variable (Dependent Variable)

Average Life Expectancy (hereafter referred to as life expectancy): It refers to the average number of years a person born in a given period can expect to live, assuming the current age-specific mortality rates (crude death rates) remain unchanged. Although it is based on current age-specific mortality rates, life expectancy is a hypothetical indicator because mortality rates are constantly changing.

2.1.2. Explanatory Variables

Table 1 outlines several categories of factors influencing population health, based on existing literature. These factors are categorized into geographic location, healthcare standards, education level, economic development, ecological environment, and urbanization level.

Table 1. Indicator Introduction.

Indicator Name	Indicator Name	Description
Geographic Location	Altitude (meters)	The altitude of provincial capitals (direct-controlled municipalities), obtained through Google Satellite Maps.
	Distance to Nearest Coastline (km)	The straight-line distance to the nearest coastline, measured using the distance tool in Baidu Maps. The starting point is the coordinates of the provincial capital, and the endpoint is the coordinates of the nearest coastline. Multiple measurements were averaged for accuracy.
Healthcare Standards	Hospital Beds per 1000 People	Data sourced from the China Statistical Yearbook and the China Health and Welfare Statistics Yearbook.
Education Level	Per Capita Education Expenditure (RMB)	Data sourced from the China Statistical Yearbook. The expenditure is calculated based on the ratio of current GDP to constant GDP, using 2000 as the base year.

Economic Development	Constant Price Per Capita GDP (RMB)	Data sourced from the China Statistical Yearbook. Calculated based on constant-price GDP for each five-year period with the base year set as 2000.
Ecological Environment	Sulfur Dioxide Emissions (10,000 tons)	Data sourced from the China Statistical Yearbook.
	Carbon Dioxide Emissions (10,000 tons)	Data sourced from the China Carbon Accounting Database.
Urbanization Level	Proportion of Urban Population (%)	Data sourced from the China Statistical Yearbook.
	Population Density (people per square kilometer)	The population of each region divided by its area. Population data is sourced from the China Statistical Yearbook.

2.2. Data Collection and Processing

2.2.1. Data Collection

After determining the indicator variables, this study selected data from 31 provincial-level administrative regions in mainland China for the years 2000, 2010, and 2020 as the research subjects. The data were sourced from the official website of the National Bureau of Statistics of China and the China Carbon Accounting Database.

2.2.2. Data Processing

During the data collection process, carbon dioxide emissions data for Tibet Autonomous Region were not available for any of the years.

A line graph was plotted for 93 data points, as shown in Figure 1, revealing that the trends in sulfur dioxide emissions and carbon dioxide emissions were similar. However, the x6 value (sulfur dioxide emissions) for Tibet in each year was close to 0, much smaller than the maximum values of other provinces. Therefore, the x7 value (carbon dioxide emissions) for Tibet in each year was estimated to be 0.

Additionally, data for Chongqing in 2000 were missing. Since Chongqing was part of Sichuan Province before 1997, the data for Sichuan Province were used as a substitute.

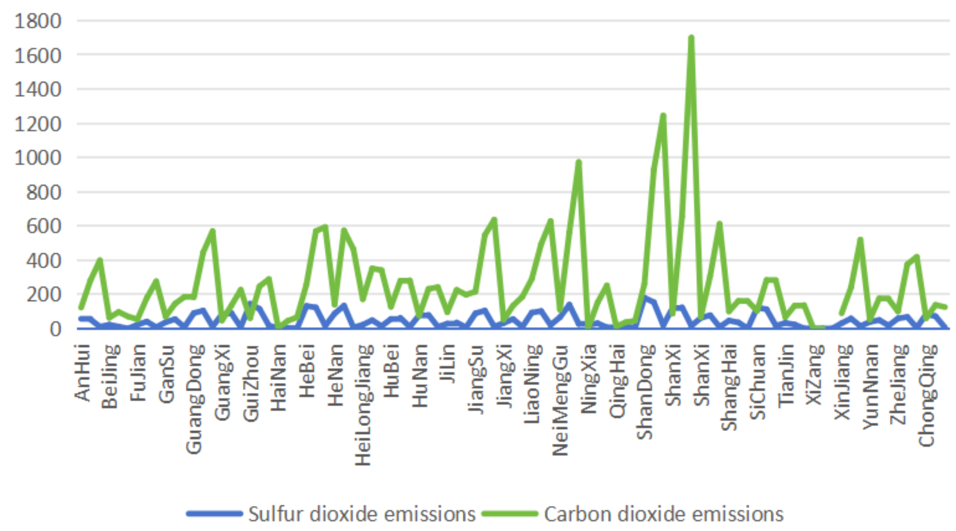


Figure 1. Emissions Line Graph.

2.3. Method Selection and Model Specification

This study employs multiple linear regression analysis to explore the relationship between the average life expectancy of the population (y) and factors such as geographical location, healthcare level, education level, economic development, ecological environment, and urbanization level ($x1, x2, \dots, x9$). Multiple linear regression analysis is a commonly used statistical method that simultaneously considers the impact of multiple independent variables on the dependent variable, and tests the significance, direction, and magnitude of each independent variable. The data processing and model fitting will be conducted using R language. It is hypothesized that the average life expectancy of the population (y) has the following linear relationship with the independent variables ($x1, x2, \dots, x9$):

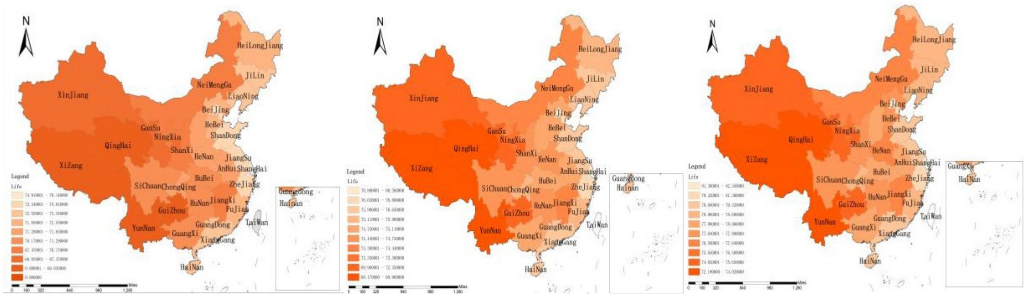
$$y = \beta 0 + \beta 1 x1 + \beta 2 x2 + \dots + \beta 9 x9 + \varepsilon$$

Where $\beta 0, \beta 1, \dots, \beta k$ are the model parameters, and ε is the error term.

The goal of this study is to estimate the regression coefficients of each variable using the least squares method, and to perform significance tests, multicollinearity tests, heteroscedasticity tests, autocorrelation tests, and outlier tests to ensure the validity and robustness of the model. Additionally, the independent variables will be standardized to facilitate comparison of the relative impact of each independent variable on the dependent variable.

3. Descriptive Statistics

The data will be observed from two perspectives: statistical measures and statistical graphics. Figure 2 shows the life expectancy heatmap of all provinces across China for the years 2000, 2010, and 2020. Figure 3 illustrates the corresponding probability distribution curves.



A. 2000 Year B. 2010 Year C. 2020 Year

Figure 2. The life expectancy heatmap of all provinces across China for the years 2000, 2010, and 2020.

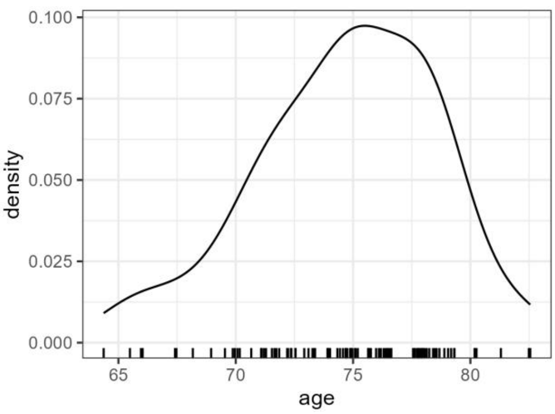


Figure 3. Probability distribution curve of life expectancy.

It can be observed that the average life expectancy data of each province over the past 20 years shows a general upward trend, increasing year by year. There is a significant difference between provinces, with the eastern regions having higher life expectancy and the western regions lower. Over time, the provincial differences have remained relatively stable. The probability distribution of life expectancy approximately follows a normal distribution, with a slight left skew.

To further examine the situation of all variables, for convenience, the variable names will be simplified (Table 2).

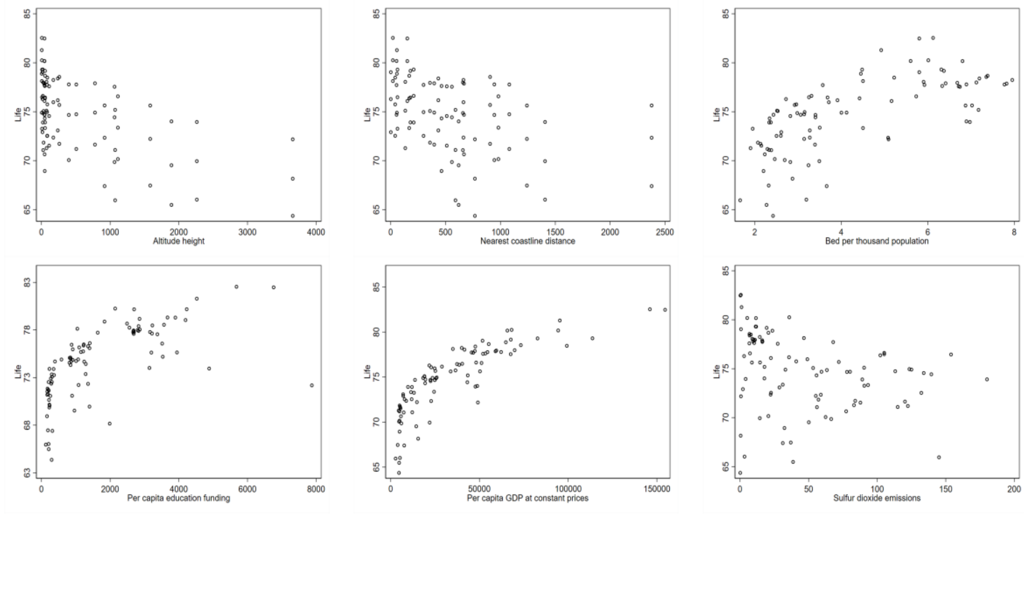
Table 2. Simplified table of variable names.

Type	Indicator Name	Simplify variable names
geographic location	Life (Year)	y
	Altitude height (m)	x1
	Nearest coastline distance (km)	x2
Characteristics of the healthcare system	Bed per thousand population	x3
Education Level	Per capita education funding (yuan)	x4
Economic Development	Per capita GDP at constant prices (yuan)	x5
Ecological Environment	Sulfur dioxide emissions (10,000 tons)	x6
	Carbon dioxide emissions (10,000 tons)	x7
Urbanization level	Proportion of urban population (%)	x8
	Population per square kilometer (people)	x9

4. Estimation Results and Explanation

4.1. Model Establishment and Testing

Figure 4 shows scatter plot of 9 independent variables relative to the dependent variable y.



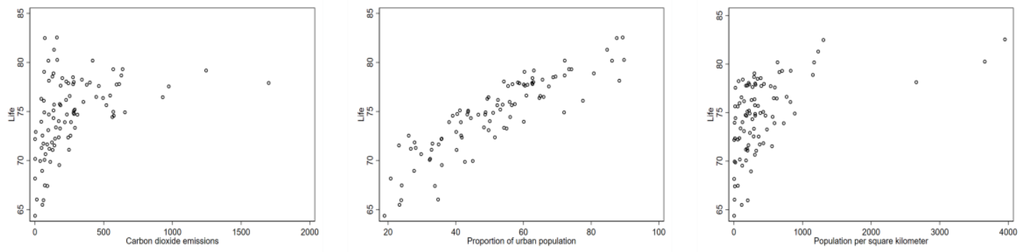


Figure 4. Scatter plot of 9 independent variables relative to the dependent variable y.

After plotting the scatter plots (Figure 5), it was found that there are certain linear or nonlinear trends between the 9 independent variables and the dependent variable. Among them, $x4$, $x5$ and $x9$ exhibited a more pronounced nonlinear relationship with the dependent variable. Therefore, these three variables were replaced with $ln(x4)$, $ln(x5+10)$ and $ln(x9)$ respectively. After the transformation, scatter plots were drawn again for each variable against the dependent variable, and the results showed a clearer linear relationship.

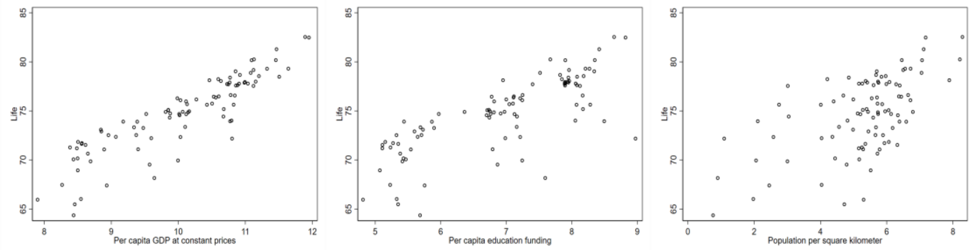


Figure 5. Scatter plots of $lnx4$, $ln\ (x5+10)$, and $lnx9$ relative to the dependent variable y.

4.2. Analysis of Influencing Factors

The F-test for the equation shows an F-statistic value of 255.2, with the corresponding p-value being approximately $2.2 \times 10^{-16} < \alpha = 0.05$. Therefore, at a significance level of 0.05, it can be concluded that the overall model is significant based on the F-test. This suggests a high correlation between the independent variables and the dependent variable y. However, the t-test results indicate that the significance of individual variables is not strong (Figure 6).

```
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 58.4626218  1.0273701  56.905  < 2e-16 ***
x1          -0.0018722  0.0001928  -9.711 2.20e-15 ***
x2          -0.0007935  0.0002406  -3.298  0.00143 **
x3           0.1468311  0.0930481   1.578  0.11832
x4           2.0192868  0.1576909  12.805  < 2e-16 ***
x6           0.0048256  0.0027211   1.773  0.07979 .
x7           0.0002112  0.0003765   0.561  0.57626
x8           0.0452014  0.0099433   4.546 1.82e-05 ***
x9           0.1848422  0.2857126   0.647  0.51943
---
Signif. codes:  0  '***'  0.001  '**'  0.01  '*'  0.05  '.'  0.1  ' '  1
Residual standard error: 0.8036 on 84 degrees of freedom
Multiple R-squared:  0.9605,    Adjusted R-squared:  0.9567
F-statistic: 255.2 on 8 and 84 DF,  p-value: < 2.2e-16
```
.

Figure 6. Direct regression results.

To improve the significance, a multicollinearity test was performed using both direct observation and the Variance Inflation Factor (VIF) method. These methods help identify potential multicollinearity issues and ensure the robustness of the regression model (Figure 7-8).

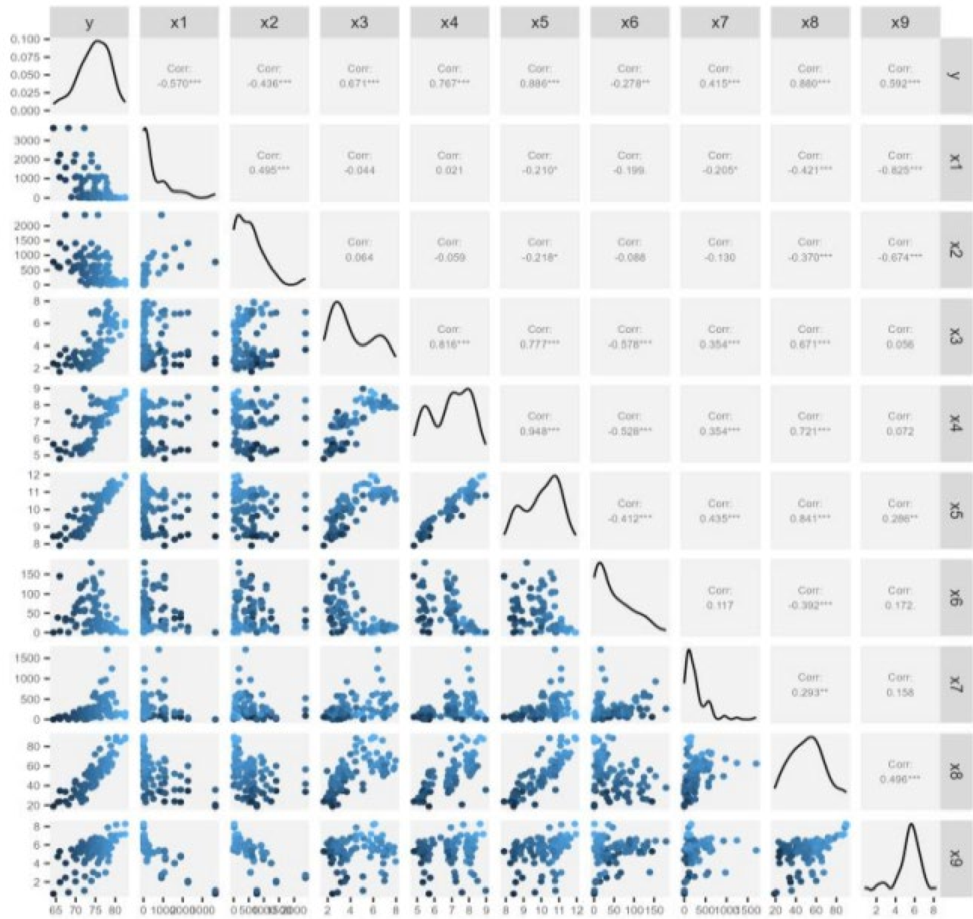


Figure 7. Visualization of correlation between variables.

```
> vif(lm1)
x1        x2        x3        x4        x5        x6        x7        x8        x9
3.992608  2.124852  4.226455  28.212447 34.644247 2.231365  1.627639  6.060691  4.890634
```

Figure 8. Direct regression multicollinearity test.

After inspection, it was found that VIF values for variables 4 and 5 were greater than 10, indicating potential multicollinearity. Meanwhile, analyze whether there are outliers in ordinary least squares.

① Regarding the outlier of the dependent variable y: Remove the student residuals |SRE|, which are all less than 3, and there are no outliers related to y.

② On the outliers of independent variables: first, calculate the leverage value of each independent variable as shown in the first two rows of the table. It is found that the 26th, 31st, 36th, 88th and 93th values, that is, the leverage value of the data points in Xinjiang, Xizang and Inner Mongolia in 2000, 2020 and 2010 is more than twice, which can be considered as the leverage point. Next, calculate the Cook distance of each independent variable as shown in the last two rows of the table, where Di is less than 1 and there are no outlier points related to the independent variable.

To eliminate multicollinearity, stepwise regression analysis method is chosen to select the most important variable from a large number of available options. Since the research question is structural analysis, we want to retain more independent variables when selecting variables. Using R, the altitude ($x1$), average coastline distance ($x2$), number of beds per thousand people ($x3$), per capita education expenditure ($x4$), per capita GDP ($x5$), sulfur dioxide emissions ($x6$), carbon dioxide emissions ($x7$), urban population proportion ($x8$), and population per square kilometer ($x9$) were used as independent variables, while the average life expectancy (y) was used as the dependent variable for stepwise regression analysis. The results are as follows in Figure 9:

```
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 55.8851444  2.1498218  25.995  < 2e-16 ***
x1          -0.0020233  0.0001821 -11.111  < 2e-16 ***
x2          -0.0004792  0.0002640  -1.815 0.073257 .
x3           0.1496370  0.0877382   1.705 0.091982 .
x4           1.4562523  0.4003266   3.638 0.000486 ***
x5           0.7053825  0.4800778   1.469 0.145671
x6           0.0041950  0.0026813   1.565 0.121637
x8           0.0416051  0.0112102   3.711 0.000379 ***
---
Signif. codes:  0  '***'  0.001  '**'  0.01  '*'  0.05  '.'  0.1  ' '  1
Residual standard error: 0.7817 on 80 degrees of freedom
Multiple R-squared:  0.9593,   Adjusted R-squared:  0.9558
F-statistic: 269.5 on 7 and 80 DF,  p-value: < 2.2e-16
Model:
y ~ x1 + x2 + x3 + x4 + x5 + x6 + x8
Df Sum of Sq     RSS     AIC
<none>                48.879 -35.743
x1      1    75.428 124.307  44.397
x2      1     2.013  50.892 -34.192
x3      1     1.777  50.656 -34.601
x4      1     8.085  56.964 -24.273
x5      1     1.319  50.198 -35.400
x6      1     1.496  50.374 -35.091
x8      1     8.416  57.295 -23.763
```

Figure 9. Gradual regression results.

It was found that x5 and x6 did not pass the *t*-test. After removing x5 with the smallest AIC value, perform stepwise regression again. The results are as follows (Figure 10):

```
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 58.8165689  0.8065150  72.927  < 2e-16 ***
x1          -0.0021301  0.0001682 -12.666  < 2e-16 ***
x2          -0.0005185  0.0002645  -1.960   0.0534 .
x3           0.1461376  0.0883310   1.654   0.1019
x4           1.9970948  0.1585030  12.600  < 2e-16 ***
x6           0.0055647  0.0025319   2.198   0.0308 *
x8           0.0498799  0.0097620   5.110 2.11e-06 ***
---
Signif. codes:  0  '***'  0.001  '**'  0.01  '*'  0.05  '.'  0.1  ' '  1
Residual standard error: 0.7872 on 81 degrees of freedom
Multiple R-squared:  0.9582,   Adjusted R-squared:  0.9551
F-statistic: 309.7 on 6 and 81 DF,  p-value: < 2.2e-16
```

Figure 10. Gradual regression results.

It was found that 6 independent variables were retained, and the *p*-value of x3 was 0.1019, which can be considered as edge significant. It can be considered as passing the significance test. Multiple collinearity tests were conducted on the new model, and the VIF values were all < 10, indicating no significant multicollinearity (Figure 11).

```
> vif(lm2_step)
x1        x2        x3        x4        x6        x8
1.997370  1.604195  3.881282  4.290668  1.684500  4.038037
```

Figure 11. Stepwise regression and multicollinearity test.

Next, we will test the basic assumptions.

1) Heteroscedasticity test

Draw a residual scatter plot with the dependent variable value as the horizontal axis as follows (Figure 12):

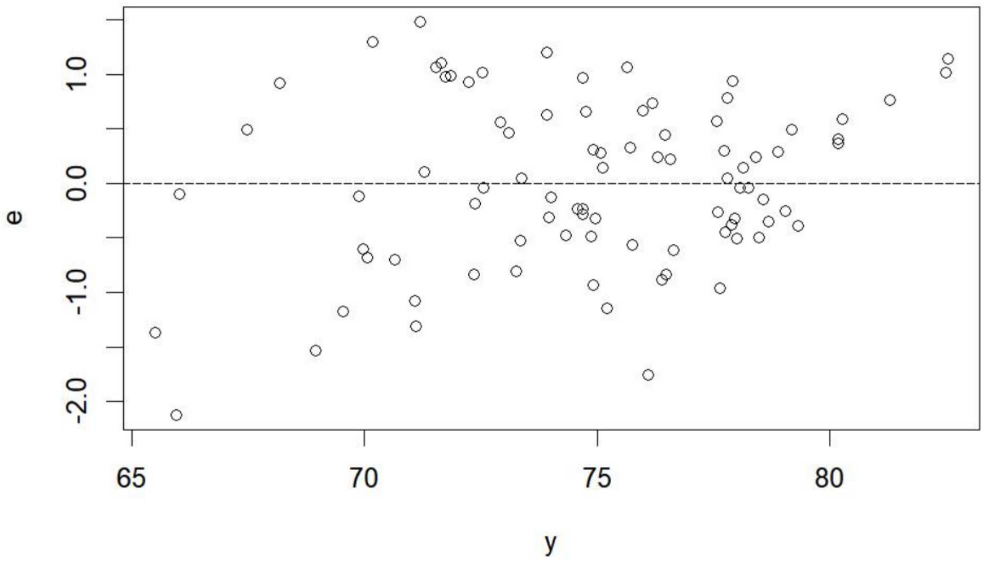


Figure 12. Residual scatter plot.

The distribution of points in the residual graph is irregular, indicating intuitively that there is no heteroscedasticity in the model.

Continuing with the Goldfeld Quanadt test, the results are as follows in Figure 13:

```
data:  lm2_step
GQ = 0.50746, df1 = 37, df2 = 37, p-value = 0.9788
alternative hypothesis: variance increases from segment 1 to 2
```

Figure 13. GQ inspection results.

GQ = 0.507, $p = 0.9788 > 0.05$, does not reject the null hypothesis, indicating that the model does not have heteroscedasticity.

2) Autocorrelation test

Perform DW inspection and obtain the following results in Figure 14:

```
lag Autocorrelation D-W Statistic p-value
1      0.03274662      1.872194    0.328
Alternative hypothesis: rho != 0
```

Figure 14. DW Result.

It can be seen that the DW value is 1.75, with a *p*-value of 0.25. According to the DW table, n = 88, k = 7, at the significance level, dL = 1.52, dU = 1.80. Since dU < DW < 4-dU, there is no autocorrelation between the error terms.

3) Outlier testing

① Regarding the outlier of the dependent variable y: Removing the studentized residual

|SRE (24)| is greater than 3, indicating that the data points in Sichuan in 2000 are outliers related to the dependent variable y.

② Firstly, calculate the leverage values of each independent variable as shown in the first two rows of the table. It was found that the leverage values of the 28th and 59th data points in Gansu in 2000 and 2010 were greater than twice, which can be considered as leverage value points. Next, calculate the Cook distance of each independent variable as shown in the last two rows of the table, where Di is less than 1 and there are no outlier points related to the independent variable.

After removing outliers, the least squares regression was performed again, and the results are shown in Table 3.

Table 3. Inspection result.

model	**Unnormalized coefficient**		**Standardization coefficient**	***t*-value**	***p*-value**
	Coefficient	**Robust**	**Beta**		
Constant	58.836	0.822	-	71.567	0.000
Altitudeheight（x1）	-0.002	0.000	-0.405	-12.727	0.000
Nearest coastline distance（x2)	-0.001	0.000	-0.067	-2.151	0.035
Bed per thousand population（x3)	0.161	0.089	0.082	1.811	0.074
Per capita education funding（x4)	1.987	0.161	0.601	12.355	0.000
Sulfur dioxide emissions（x6)	0.006	0.003	0.069	2.248	0.027
Proportion of urban population（x8)	0.050	0.010	0.234	5.107	0.000

Dependent variable: life(Y).

Model summary shows in Table 4.

Table 4. Model summary.

Model	**R**	**R^2**	**Ad-R^2**	**Error in standard estimates**
1	0.978[a]	0.957	0.954	0.787

Predictive variable:(Constant), x8, x1, x6, x2, x3, x4.

The p-value of x3 is 0.704, indicating edge saliency. It can be considered as passing the significance test.

Based on the above analysis, the final model formula is:

$$\hat{y} = 58.836 - 0.002x1 - 0.001x2 + 0.161x3 + 1.987x4 + 0.006x6 + 0.050x8$$

Where $x1$ represents altitude, $x2$ represents average coastline distance, $x3$ represents number of beds per thousand population, $x4$ represents per capita education expenditure, $x6$ represents sulfur dioxide emissions, and $x8$ represents the proportion of urban population.

Due to the different units of measurement and forms of each independent variable, there are significant differences in data size. Therefore, standardized regression coefficients are calculated to compare the relative importance of each variable's impact on y. The formula is as follows:

$$\hat{y} *= -0.405x1 * -0.067x2 * +0.082x3 * +0.601x4 * +0.069x6 * +0.234x8 *$$

4.3. Model Analysis and Research Conclusions

The estimation model presented above primarily analyzes the relationship between factors such as altitude, proximity to the coastline, bed availability per thousand people, per capita education expenditure, sulfur dioxide emissions, and the proportion of the urban population, and life expectancy across various regions in China.

The results from the model show that for every 1-meter increase in altitude and every 1-kilometer increase in the distance to the nearest coastline, life expectancy decreases by approximately 0.2% and 0.1%, respectively. Both of these factors exhibit a negative correlation with life expectancy, indicating that coastal and low-altitude areas tend to have higher life expectancy compared to inland, high-altitude regions.

Furthermore, for every additional bed per thousand people, life expectancy increases by 8.2%, highlighting the significant impact of healthcare conditions on life expectancy. This result underscores the importance of improving healthcare infrastructure in promoting longer life expectancy.

Additionally, for every 1 yuan increase in per capita education expenditure, life expectancy rises by 198%. A substantial body of literature has emphasized the role of education in enhancing individuals' ability to acquire knowledge, solve problems, and exercise self-control [4]. These factors ultimately lead to healthier lifestyle choices, with increased education levels fostering a greater awareness of personal health and health-related knowledge, thus contributing to an increase in life expectancy.

An unexpected result was observed with sulfur dioxide emissions. For every unit increase in sulfur dioxide emissions, life expectancy increases by 0.6%. This positive correlation contradicts the expected negative impact of pollution on life expectancy. The potential explanation for this outcome lies in the model's limitations, particularly with respect to endogeneity issues that were not addressed. There may be reverse causality or omitted variable bias at play. It is possible that pollution emissions, while typically harmful, could be associated with higher levels of industrialization or urbanization, which may indirectly lead to improvements in healthcare and nutritional standards, thereby potentially affecting life expectancy. Further research and control of relevant variables, involving more specialized sociological and medical knowledge, are needed to resolve this issue.

In terms of urbanization, for every 1% increase in the proportion of the urban population, life expectancy increases by 5%, showing a positive correlation. A higher urbanization rate generally indicates better access to social security and healthcare services, which can mitigate the impact of external factors that reduce life expectancy. Moreover, higher urbanization typically corresponds to improved living conditions, which help reduce endogenous mortality, thereby enhancing life expectancy [5].

To provide a more intuitive understanding of the relative importance of each influencing factor, the regression model was standardized. The resulting coefficients were as follows: Altitude: -0.405; Average coastline distance: -0.067; Bed availability per thousand people: -0.082; Per capita education expenditure: 0.601; Sulfur dioxide emissions: 0.069; Urban population proportion: 0.234.

From these results, it can be concluded that per capita education expenditure and the urban population proportion have the most significant impact on life expectancy, while the other factors have a relatively smaller effect.

5. Policy Implications and Directions for Future Research

Based on the conclusions drawn from the estimation model, several policy implications can be derived.

5.1. Strengthening Investment in Education and its Accessibility

Per capita education expenditure is one of the most significant factors influencing life expectancy, suggesting that education plays a critical role in improving public health and quality of life. Therefore, it is recommended that the government increase investment in education, enhance the quality and coverage of education, particularly in poverty-stricken and rural areas. Efforts should focus on strengthening both basic and vocational education, improving cultural literacy and skill levels among the population, and enhancing their awareness of health knowledge and self-care capabilities.

5.2. Optimizing Medical Resources and Their Distribution

The number of hospital beds per thousand people is a major determinant of life expectancy, indicating that healthcare conditions are crucial for disease prevention and treatment. It is thus suggested that the government optimize the allocation of medical resources, strengthen the primary healthcare system, and reduce the healthcare disparities between urban and rural areas, as well as across different regions. This would improve the quality and efficiency of healthcare services, reduce healthcare costs, and ensure that all people have access to equal, convenient, efficient, and safe medical care.

5.3. Promoting Urbanization and the New-Type Urbanization Process

The proportion of the urban population is another important factor influencing life expectancy, showing that urbanization plays a key role in improving living conditions and the environment. Therefore, it is recommended that the government accelerate the urbanization process, particularly by facilitating the integration of migrant populations into cities, enhancing the urban-rural social security system, and improving income and consumption levels of both urban and rural residents. Additionally, efforts should focus on advancing new-type urbanization, optimizing urban planning and management, improving urban functions and quality, and creating modern cities that are livable, business-friendly, and tourism-oriented.

5.4. Strengthening Ecological Environmental Protection and Governance

Sulfur dioxide emissions were found to be positively correlated with life expectancy, which contradicts the expected negative correlation with pollution levels. This anomaly may be due to model limitations and unresolved endogeneity issues. In fact, numerous studies have shown that pollutant emissions have a negative impact on public health and life expectancy. Therefore, it is recommended that the government intensify ecological environmental protection and governance, adhere to a green development philosophy, implement energy-saving and emission-reduction targets, and increase regulation and enforcement in key industries and regions. Efforts should be made to reduce pollutant emissions, improve air and water quality, and ensure a good ecological environment for the public.

5.5. Policy Formulation Should Be Context-Specific, Taking Geographic Factors into Account

Altitude and proximity to the coastline are relatively minor factors influencing life expectancy, suggesting that geographical factors also have some impact. Therefore, when formulating and implementing relevant policies, it is important to take into account the geographical context, including the natural conditions, resource endowments, and economic development levels of different regions. Tailored and differentiated policies should be developed to avoid a "one-size-fits-all" approach, ensuring policies are applied based on the specific circumstances of each area.

5.6. Limitations

Several limitations in this study must be addressed for more refined conclusions.
First, in the selection of independent variables, nutrition, health, and lifestyle habits are also important factors affecting public health and life expectancy. However, due to data limitations, these factors could not be incorporated into the analysis.

Second, the model formulation has certain limitations, particularly concerning unresolved endogeneity issues. For instance, reverse causality or omitted variable bias may explain why sulfur dioxide emissions show a positive correlation with life expectancy, contrary to expectations. It is crucial to control for relevant variables, and addressing these issues would require more specialized sociological and medical knowledge for further research. Future studies should employ more representative data for verification, as data availability permits.

Third, while the model explains the influence of certain factors on life expectancy, the factors affecting residents' savings rates are not limited to the variables selected in this study. Moreover, the effects of important variables, such as per capita GDP, could not be adequately explained due to difficulties in quantifying these variables or identifying appropriate models for fitting. This leaves out some key factors, and potential endogeneity issues may remain.

Finally, the conclusions derived from the estimation model reflect the relationships between life expectancy and the selected explanatory variables only for the specific data period analyzed. Given that development phases may vary in the past or future, the model may change over time, leading to different conclusions depending on the stage of development under study.

6. Conclusion

In conclusion, this study highlights the complex relationship between life expectancy and various regional factors in China. Multiple linear regression analysis reveals that education expenditure, urban population proportion, healthcare resources, and environmental factors significantly impact life expectancy. The findings emphasize the need for targeted policy interventions to enhance healthcare access, educational opportunities, and urbanization. Furthermore, addressing ecological issues and tailoring policies to regional contexts will be crucial in improving life expectancy across China. Future research should consider additional factors such as lifestyle habits and address endogeneity concerns for a more comprehensive understanding of life expectancy determinants.

References

1. K. J. Foreman, N. Marquez, A. Dolgert, K. Fukutaki, N. Fullman, M. McGaughey, et al., "Forecasting life expectancy, years of life lost, and all-cause and cause-specific mortality for 250 causes of death: reference and alternative scenarios for 2016–40 for 195 countries and territories," *The Lancet*, vol. 392, no. 10159, pp. 2052-2090, 2018, doi: 10.1016/S0140-6736(18)31694-5.
2. M. H. Vik and E. Carlquist, "Measuring subjective well-being for policy purposes: The example of well-being indicators in the WHO 'Health 2020' framework," *Scand. J. Public Health*, vol. 46, no. 2, pp. 279-286, 2018, doi: 10.1177/1403494817724952.
3. C. S. Vogeler, A. van den Dool, and M. Chen, "Programmatic action in Chinese health policy—The making and design of 'Healthy China 2030'," *Rev. Policy Res.*, vol. 40, no. 4, pp. 534-552, 2023, doi: 10.1111/ropr.12533.
4. L. Gu, M. Rosenberg, L. Yang, J. Yu, and B. Wei, "A spatial multilevel analysis of the impacts of housing conditions on county-level life expectancy at birth in China," *Appl. Geogr.*, vol. 124, p. 102311, 2020, doi: 10.1016/j.apgeog.2020.102311.
5. S. Bandyopadhyay and E. Green, "Urbanization and mortality decline," *J. Reg. Sci.*, vol. 58, no. 2, pp. 483-503, 2018, doi: 10.1111/jors.12375.