Research on Spatial Autocorrelation of the Population in Shandong Province

Silian Shen[1, a], Xinqian Wu[2, b], Chunwei Wang[3, c]

[1]School of Mathematics and Statistics, Henan University of Science and Technology, Luoyang, 471023, China

[2]School of Mathematics and Statistics, Henan University of Science and Technology, Luoyang, 471023, China

[3]School of Mathematics and Statistics, Henan University of Science and Technology, Luoyang, 471023, China

[a]shensilian@126.com, [b]cuipersuit@126.com, [c]wangpersuit@126.com

Keywords: Spatial Autocorrelation; Moran I Statistics; Visualization technique

Abstract. Spatial autocorrelation is one of the most important characteristics of spatial data. Based on the fifth and sixth census data, we focus on in this paper studying the spatial autocorrelation of the population in Shandong province. Specifically, the global Moran I statistic is first used to quantify the spatial clustering of the population in the whole province both in 2000 and 2010 years, respectively. Then, the local Moran I statistic is employed to assess the spatial autocorrelation degree between one city and others. Finally, related results are visualized by applying the visualization technique of the Surfer software.

Introduction

With the rapid growth of population in the world in recent decades, the problem of population is becoming more and more important. China is the most populous country in the world. Shandong is the second populous province in China. Therefore, it is necessary to study the spatial autocorrelation in order to capture more detailed features of the population among the countries and cities in Shandong province.

Recently, some new models and technologies are employed to unveil the patterns of the population distribution (see [1-11] for more details). For example, Li and Feng et al. used the logistic model (see the literature [1, 2] for details) to study the total population all over the mainland of China. Linear regression and auto-regression methods are taken to analyze the development patterns of the population in China (see [6-8]). Based on the grey model, Xue and Zhou investigated the growth state of the population in China and predicted the total population in the future several years ([9, 10]). Furthermore, the neural network is used to predict the total population in Xiangfan city ([11]).

Although these models mentioned in the above literature can explore the features of the population to some extent, most of them are global numeric indicators which are lack of details and intuitions. Therefore, there is still considerable space for research with respect to the spatial distribution of the population, especially for spatial autocorrelation.

Based on the fifth and sixth census data which is from the statistical yearbook of Shandong province, we focus on in this paper exploring the spatial autocorrelation of the population. Specifically, the spatial autocorrelation degree of the population is first investigated throughout the whole province by using the global Moran I statistic. Then, the local Moran I coefficient is employed to evaluate the local spatial autocorrelation. Finally, related results are visualized by applying the visualization technique of the Surfer software.

Data Sources

In this section, we choose the total population data of 132 countries and cities in Shandong

province. The spatial distribution of these locations is shown in Figure 1.   By using the Google Earth software, the spatial coordinates, saying the latitude and longitude data (unit: degree), of each location are collected. Furthermore, related population data in 2000 year is also taken so that some comparisons are conducted.
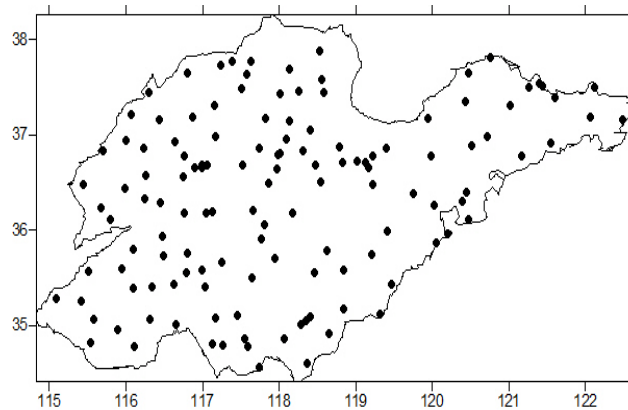

Figure 1. Spatial distribution of 132 countries and cities in Shandong province.

In can be seen from Figure 1 that, on one hand, the selected spatial locations are almost uniformly distributed throughout the whole province. On the other hand, all the locations (denoted by dot) are within the borders of Shandong province, which demonstrates that the latitude and longitude data collected by Google Earth software is comparatively accurate. The above two points are also necessary for us to perform valid analysis and reasonable explanation.

Global Spatial Autocorrelation of the Population

In this section, the global Moran I statistic is first used to quantify the spatial clustering patterns of the population throughout the whole province. Then, the local Moran I coefficients are applied to examine whether the total population of certain location is consistent with the corresponding population of its surrounding ones.

In statistics, the global Moran I coefficient is commonly used to assess the autocorrelation among all locations or observations. Along this line of thinking, we extend this coefficient to quantify the spatial autocorrelation of the population. Specifically, we define the spatial global Moran I statistics as follows:

$$I = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \overline{x})(x_j - \overline{x}) \Big/ \left(S^2 \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} \right), \tag{1}$$

where

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \text{ and } S = \left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 \right)^{\frac{1}{2}} \tag{2}$$

are the sample mean and sample standard deviation of the population data. x_i and x_j stand for the total population of the ith location and that of the jth location, respectively. w_{ij} denotes the link relationship between the ith location and the jth location, which can be determined according to the adjacent criterion or distance criterion. Specifically, the former is described as:

$$w_{ij} = \begin{cases} 1, & \text{if location } i \text{ and location } j \text{ are adjacent,} \\ 0, & \text{if location } i \text{ and location } j \text{ are not adjacent,} \end{cases} \tag{3}$$

where $i \neq j$.

The latter rule is expressed as:

$$w_{ij} = \begin{cases} 1, & \text{if the distance beteewn the location } i \text{ and location } j \text{ is smaller or equal to } d, \\ 0, & \text{if the distance beteewn the location } i \text{ and location } j \text{ is greater than } d, \end{cases} \tag{4}$$

where $i = 1, 2, L, n$; $j = 1, 2L, n$ and $i \neq j$, d is a prespecified constant. It should be noted that the adjacent criterion is adopted in this paper and a binary link matrix is obtained.

With the data and the spatial global Moran I formula described in equation (1), the corresponding values both in 2000 and 2010 years are 0.1947669 and 0.1664098, respectively. From the results, we can observe that, both the coefficient values are significantly greater than the expectation of the corresponding approximate normal distribution, which implies that the population in Shandong province has strong spatial autocorrelation and this province has good population structure. Furthermore, the value of the Moran I coefficient in 2010 year is slightly smaller than that in 2000 year, which indicates that the spatial autocorrelation of the population is decreasing. This may be caused by the rapid development of economic and society.

Local Spatial Autocorrelation of the Population

In this section, the local Moran I statistic is extended to assess and evaluate the spatial autocorrelation locally. Specifically, for each of $i = 1, 2, L, n$, it is described as

$$I_i = (x_i - \overline{x}) \sum_{j=1}^{n} w_{ij} (x_j - \overline{x}) \bigg/ S^2 ,$$ (5)

where $\overline{x}, S^2$ and w_{ij} are shown in equations (2) and (3).

Based on the fifth and sixth census data in Shandong province, the local Moran I coefficient values of the 132 countries and cities are computed to quantify the local spatial autocorrelation of the population. In order to demonstrate the detailed patterns of the spatial autocorrelation between one location and its surrounding ones intuitively, we use the Surfer software to visualize the coefficient values. Related results are shown in contour lines and depicted in Figure 2.
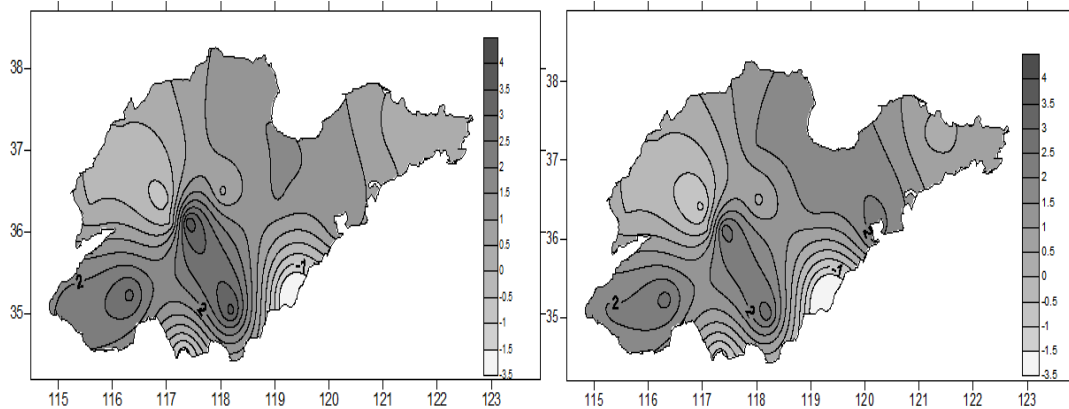

Figure 2. Contour maps of the local Moran I coefficient values of 132 cities in Shandong province: left: 2000 year and right: 2010 year.

It can be seen from Figure 2 (left) that: (1) The counties and cities whose Moran I correlation coefficient values are greater than 1.3 account for 47.06% while the population in these corresponding regions makes up 53.23% of the whole province's population. This means that most of the counties and cities are strongly spatial autocorrelated and similar, especially for Qingdao, Dongying, Weifang, Jining, Laiwu, Linyi, Binzhou and Heze. (2) The regions whose coefficient values are between 0 and 1.3 account for 29.41% and the total population in these regions makes up 27.03%, which implies that these regions have weakly spatial autocorrelation and internal similarity. (3) The regions whose coefficient values are negative occupies 23.53% while the total population in these corresponding locations accounts for 19.61%, which demonstrates that these locations are significantly dissimilar and the larger internal deviation is present.

From the results in Figure 2 (right), we can observe that the variation of the spatial autocorrelation of the population displays similar patterns to that in 2010 year. These patterns are omitted here to save space.

Conclusion

Based on the data which is from the statistical yearbook, the spatial autocorrelation of the total population in Shandong province is investigated in this paper. Specifically, the global and local Moran I coefficients are extended to assess and evaluate the spatial dependence of the population in distinct regions globally and locally. And related results are visualized by using the Surfer software.

References

[1] Baisui Li, Liga Tong. The Logistic Model of Population Urbanization in Inner Mongolia and Application [J]. Journal of Arid Land Resources and Environment. 2007 (2) 32-36.

[2] Shouping Feng. On Prediction Model of China Population Growth [J]. Journal of Anhui Science and Technology University. 2008 (6) 73-76.

[3]　Wenquan Chen and so on. The Application of Leslie Revised Model in Population Prediction [J]. World Science-Technology Research and Development. 2008 (2) 219-224.

[4] Nan Li and so on. The Prediction of Chinese Stochastic Population [J]. Chinese Journal of Population Science. 1998 (1) 10-16.

[5] Nan Li and so on. The Population Forecasting and Project Based on Stochastic Model [J]. Systems Engineering—Theory Methodology Applications. 1998 (1) 37-41.

[6] Ying Fu. Application of Regression Analysis to Population Prediction [J]. Liaoning Higher Vocational Technical Institute Journal. 2000 (1) 56-58.

[7] Xundong Li. Study on the Natural Environmental Factors Affecting Population Distribution in the Guizhou Karst Plateau: Multivariate Regression Analysis and Zonality [J]. Arid Zone Research. 2007 (2) 280-286.

[8] Ai Chen. An. A Study on Chinese Population Forecast by Time Series Model [J]. Population and Economics. 2004 (6) 63-67.

[9] Zhen Xue. Mathematical Model for Chinese Population Growth Forecast [J]. Journal of Henan Institute of Science and Technology. 2008 (1) 123-127.

[10] Shouping Zhou. Study on the Grey Predictive Model and Its Application to Chinese Population [J]. Journal of Mathematical Medicine. 2005 (4) 307-309.

[11] Jing Zhang and Xinhua Wang. Forecasting the Amounts of Population Based on Artificial Neural Network [J]. Journal of Xiangfan University. 2001 (5) 73-76.