

基于相似地理加权回归模型的中国区域人口死亡率影响因素空间异质性分析

贾丰铭, 冯文羽, 郑天琦, 魏传华*

中央民族大学理学院统计学系, 北京

收稿日期: 2025年6月22日; 录用日期: 2025年7月12日; 发布日期: 2025年7月25日

摘要

人口死亡率与其影响因素之间的空间异质性是人口学和管理学等多个领域关注的重点。地理加权回归模型经常被用来处理地理数据中的空间异质性问题, 然而该方法作为一种局部建模方法, 只考虑了观测个体之间空间距离上的远近关系, 没有考虑观测个体之间属性相似性。为了解决这一问题, 本文使用相似地理加权回归模型, 从人口结构、社会经济、医疗条件三个方面, 分析2020年中国区域人口死亡率的空间格局及其影响机制。该模型在构造回归系数的局部加权最小二乘估计时, 权函数同时考虑了地理邻近性与属性相似性, 拟合效果更好。结果表明, 人口结构、社会经济、医疗条件的解释变量对人口死亡率的影响随空间位置变化, 具有明显的空间异质性, 且各因素对于各地级行政区的人口死亡率具有不同程度和方向的影响。针对空间异质性特征, 探讨了模型结果的稳健性, 并提出了差异化的区域政策建议。

关键词

空间异质性, 相似地理加权回归模型, 人口死亡率

Spatial Heterogeneity Analysis of Influencing Factors for Regional Population Mortality Rates in China Based on the Similarity and Geographically Weighted Regression Models

Fengming Jia, Wenyu Feng, Tianqi Zheng, Chuanhua Wei*

Department of Statistics, School of Science, Minzu University of China, Beijing

Received: Jun. 22nd, 2025; accepted: Jul. 12th, 2025; published: Jul. 25th, 2025

*通讯作者。

文章引用: 贾丰铭, 冯文羽, 郑天琦, 魏传华. 基于相似地理加权回归模型的中国区域人口死亡率影响因素空间异质性分析[J]. 统计学与应用, 2025, 14(7): 210-220. DOI: 10.12677/sa.2025.147198

Abstract

Spatial heterogeneity between population mortality rates and their influencing factors is a key focus in fields such as demography and management science. Geographically Weighted Regression (GWR) models are frequently used to address spatial heterogeneity. However, as a local modeling technique, GWR only considers the spatial proximity between observation points and ignores their attribute similarity. To address this limitation, this study employs a Similarity-Geographically Weighted Regression (SGWR) model. Using data from three dimensions—population structure, socio-economic factors, and healthcare conditions—it analyzes the spatial patterns and influencing mechanisms of regional population mortality rates in China in 2020. This model constructs locally weighted least squares estimates for regression coefficients, incorporating both geographical proximity and attribute similarity into the weighting function, resulting in improved model fit. The results demonstrate that the influence of explanatory variables from population structure, socio-economic factors, and healthcare conditions on population mortality varies significantly across spatial locations, exhibiting clear spatial heterogeneity. Furthermore, each factor exerts varying degrees and directions of influence on mortality rates across different prefecture-level administrative divisions. Considering these spatial heterogeneity characteristics, the study examines the robustness of the model results and proposes tailored regional policy recommendations.

Keywords

Spatial Heterogeneity, Similarity and Geographically Weighted Regression Model, Population Mortality Rate

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

人口死亡率常作为人群健康水平的衡量指标，人口死亡率及其影响因素的构成从侧面反映出一个国家的社会环境发展水平和自然环境现状[1]。Brenner (1979) [2]提出，经济发展水平较低将导致人群的健康状况恶化，从而导致人口死亡率上升。中国是发展中国家，且人口基数较大。自 1978 年至 2019 年，我国的人口死亡率由 6.25‰ 上升至 7.09‰，同时人口出生率由 18.25‰ 下降至 10.41‰，自 2020 年至 2023 年，受新冠疫情等多方面因素的影响，我国人口死亡率由 7.07‰ 上升至 7.87‰，同时人口出生率由 8.52‰ 下降至 6.39‰，人口自然增长率已出现负数。因此，当前我国的人口增长模式出现负增长态势，且面对着较为严峻的人口老龄化问题，研究我国的人口死亡率及其影响因素有较大的意义。

2. 相关文献综述

目前国内外关于人口死亡率的研究主要聚焦于死亡率的预测及其影响因素的探究等方向。Lee 和 Carter (1992) [3]通过对美国人口死亡率数据的长期趋势进行分析，提出了年龄别死亡率的对数计算公式——Lee-Carter 模型。由于具备较好的实用性，Lee-Carter 模型被广泛地运用到许多国家和地区的人口死亡率预测上。陶祥兴等(2022) [4]通过分数布朗运动驱动的 O-U 过程进行 Lee-Carter 模型的拟合和预测，并引入了机器学习与时间序列相结合的 ARIMA-LSTM 方法对死亡率的预测进行改进，证明了机器学习方法对改进死亡率预测的可行性。在对于人口死亡率影响因素构成的研究中，诸多国内外成果表明，人

口死亡率与人口发展水平、社会环境等因素紧密相连。Hardarson 等(2001) [5]探究了人口死亡率与居民教育水平的相关关系。Jusot (2006) [6]基于收入视角对人口死亡率及其影响因素进行研究。赵锦辉(1994) [7]表明人口死亡率的影响因素包含社会经济发展水平、文化教育程度和医疗卫生保健条件等。

目前,学者对于人口死亡率的研究主要集中在全国或单一地区的整体分析,就我国的国情而言,省际结构差异及发展不均衡要求我们如果要探究我国的人口死亡率及其影响因素,需要考虑地区差异、影响因素的空间异质性,对我国的人口死亡率进行综合分析。应奎等(2022) [8]通过变异系数和空间自相关分析等方法探究中国人口死亡率的空间格局差异及演变趋势,发现中国各省人口死亡率的变异系数存在较大差异,反映了我国人口死亡率存在明显的空间异质性。

空间异质性可以通过空间上变化的属性值和生成这些属性数据的地理过程来观察。Brunsdon 等(1996) [9]提出的地理加权回归模型(Geographically weighted regression, GWR)常用于处理地理数据中的空间异质性问题。此模型的核心思想是在回归分析中引入空间权重,以反映地理位置对回归关系的影响。目前国内基于 GWR 模型对人口死亡率进行研究的成果数量相对较少。苏晓丽(2022) [10]发现我国五岁以下儿童的死亡率具有明显的地区分布和空间集聚特征,并以人口因素、社会经济因素和环境因素构建五岁以下儿童死亡率的时空地理加权回归模型,探究各影响因素对我国五岁以下儿童死亡率的异质性影响。赵宇等(2019) [11]在 GWR 模型的基础上,提出了一类地理加权 CBD 模型的空间变系数死亡率预测模型,进一步揭示了人口死亡率数据在空间意义上的相互联系。

先前关于 GWR 模型的研究都以地理邻近性作为创建权重矩阵的决定因素。虽然地理邻近性是 GWR 模型中的一个关键因素,但它并不一定意味着其他相关属性的一致性。Zhu 和 Turner (2022) [12]强调,地理学第三定律本质上说明,如果两个地点具有相似的地理条件,那么它们可能会针对特定的地理变量表现出类似的结果,即相似性不仅限于地理邻近性,还扩展到地理条件和数据中其他属性的排列和层次结构。因此,地理上的距离未必能准确代表实际的相似程度。对此, Lessani 和 Li (2024) [13]提出了一种在局部回归模型中结合地理邻近性和属性相似性的新方法,称为相似地理加权回归模型(Similarity and geographically weighted regression, SGWR)。这种方法旨在捕捉空间邻近性和属性相似性之间复杂的相互作用,以便更深入地理解空间模式。

本文选取相似地理加权回归(SGWR)模型,结合地理邻近性和属性相似性对中国地级行政区人口死亡率的空间异质性进行分析,探究人口死亡率与人口结构、社会经济与医疗条件等影响因素之间的关系。本文将地理学领域中的新方法引入人口死亡率的研究当中,使得人口死亡率的空间异质性研究不再局限于物理距离上,填补了人口死亡率研究中的空白。

3. 模型介绍

(一) 地理加权回归模型

地理加权回归模型(GWR)通过基于地理位置进行局部回归分析,捕捉空间数据集中的地理变化和异质性。在传统的回归分析中,通常假设模型的参数是全局的,即在整个数据集中是一致的。然而,在某些情况下,数据集在空间上的分布是不均匀的,且其分布规律可能随着空间位置的变化而变化。GWR 在空间上的每个位置分别估计回归参数,这意味着对于每个地理位置,都会有一个特定的回归系数,因此, GWR 模型可以更好地捕捉空间数据中的地理差异和变化。

GWR 模型的一般形式如下:

$$y_i = \beta_0(u_i, v_i) + \sum_{j=1}^p \beta_j(u_i, v_i) x_{ij} + \varepsilon_i, i = 1, 2, \dots, n \quad (1)$$

一般采用局部加权最小二乘法对 GWR 模型进行估计。具体来讲,对于研究区域内的一个给定的位

置 (u_0, v_0) , d_{0i} 是点 (u_0, v_0) 与第 i 个观测位置 (u_i, v_i) 之间的距离, 通常取欧式距离:

$$d_{0i} = \sqrt{(u_0 - u_i)^2 + (v_0 - v_i)^2} \quad (2)$$

则点 (u_0, v_0) 处的各系数函数 $\beta_j(u_0, v_0)$ ($j = 0, 1, 2, \dots, p$) 的局部估计可以通过使以下目标函数达到最小来得到:

$$\sum_{i=1}^n \left[y_i - \beta_0(u_0, v_0) - \sum_{j=1}^p \beta_j(u_0, v_0) x_{ij} \right]^2 K_h(d_{0i}) \quad (3)$$

其中 $K_h(\cdot) = K(\cdot/h)$ 是核函数, 一般取高斯核函数, h 是光滑参数也就是窗宽。

令

$$X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix}$$

$$Y = (y_1, y_2, \dots, y_n)^T, \quad \varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T$$

$$\beta(u_0, v_0) = [\beta_0(u_0, v_0), \beta_1(u_0, v_0), \dots, \beta_p(u_0, v_0)]^T$$

$$W_G(u_0, v_0) = \begin{pmatrix} K(d_{00}/h) & & & \\ & K(d_{01}/h) & & \\ & & \ddots & \\ & & & K(d_{0n}/h) \end{pmatrix} \quad (4)$$

对目标函数进行求解, 得到 $\beta_0(u_0, v_0), \beta_1(u_0, v_0), \dots, \beta_p(u_0, v_0)$ 的估计为

$$\hat{\beta}(u_0, v_0) = [X^T W_G(u_0, v_0) X]^{-1} X^T W_G(u_0, v_0) Y \quad (5)$$

将点 (u_0, v_0) 推广到点 $(u_1, v_1), (u_2, v_2), \dots, (u_n, v_n)$, 从而得到 $\beta(u_i, v_i)$ 的估计

$$\hat{\beta}(u_i, v_i) = [X^T W_G(u_i, v_i) X]^{-1} X^T W_G(u_i, v_i) Y \quad (6)$$

则 y_i 的拟合值为

$$y_i = x_i^T \hat{\beta}(u_i, v_i) = x_i^T [X^T W_G(u_i, v_i) X]^{-1} X^T W_G(u_i, v_i) Y \quad (7)$$

(二) 相似地理加权回归模型

GWR 模型以地理位置距离构建权重矩阵, 只考虑了地理邻近性的“邻近”概念, 而忽略了基于属性相似性的“相关”概念(Anselin 2020 [14])。相似地理加权回归模型(SGWR)在 GWR 模型的基础上, 通过构建由属性相似性和地理邻近性共同作用的权重矩阵来解决 GWR 模型中的这一缺陷。因此, SGWR 模型能够更准确地捕获跨空间的空间关系和依赖性。

SGWR 模型可以认为是具有以下形式的空间变系数模型:

$$y_i = \beta_0(u_i, v_i, x_i) + \sum_{j=1}^p \beta_j(u_i, v_i, x_i) x_{ij} + \varepsilon_i, i = 1, 2, \dots, n \quad (8)$$

其中, $(y_i; x_{i1}, x_{i2}, \dots, x_{ip})$ 是在数据点 (u_i, v_i, x_i) ($i = 1, 2, \dots, n$) 处因变量 y 和自变量 x_1, x_2, \dots, x_p 的观测值, $\beta_j(u_i, v_i, x_i)$ ($j = 0, 1, 2, \dots, p$) 是第 i 个数据点 (u_i, v_i, x_i) 处的未知参数, $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ 为独立同分布的误差项,

通常假定均值为 0，方差为 σ^2 。

与 GWR 模型类似，关于 SGWR 模型中的参数估计，采用局部加权最小二乘估计。具体来讲，对于研究区域内的一个给定的位置和其对应属性值 (u_0, v_0, x_0) ，点 (u_0, v_0, x_0) 处的各系数函数 $\beta_j(u_0, v_0, x_0) (j = 0, 1, 2, \dots, p)$ 的局部估计可以通过使以下目标函数达到最小来得到

$$\sum_{i=1}^n \left[y_i - \beta_0(u_0, v_0, x_0) - \sum_{j=1}^p \beta_j(u_0, v_0, x_0) x_{ij} \right]^2 w_i(u_0, v_0, x_0) \quad (9)$$

其中，SGWR 模型的权重矩阵 $W(u_0, v_0, x_0)$ 由地理权重和属性权重两部分构成。地理权重部分与 GWR 模型相同，如公式(4)所示，即基于 Tobler 第一定律思想，离数据点 (u_0, v_0) 越近的对参数估计的贡献越大。

SGWR 模型的属性权重本质是根据自变量来计算回归位置与其他观测值之间的属性相似度，为与回归点相似度较高的邻居分配较高的权重，与回归点相似度较低的邻居分配较低的权重。计算回归位置 i 和观测值 j 之间成对距离的公式如下：

$$d(i, j) = \frac{1}{p} \sum_{k=1}^p |x_{k(i)} - x_{k(j)}| \quad (10)$$

属性的成对距离值越低表示相似性越高，然而，SGWR 模型中的地理权重被设计为越接近 1 的值代表观测点距离回归位置越近，而接近 0 则认为距离更远。因此，需要将属性权重与地理权重的变化趋势保持一致，即接近 1 的值表示两个位置的属性之间有较高的相似性，接近 0 的值表示较低的相似性。公式如下：

$$W_S(x_i) = e^{-d(i,j)^2} \quad (11)$$

在计算属性相似性的成对距离之前，需对数据进行标准化，确保每个变量的均值为零，标准差为 1。这种标准化过程抵消了变量之间不同大小的影响，从而可以在各位置之间进行更准确且与尺度无关的相似性测量。

将地理权重与属性权重按照一定的贡献比例相结合，构成 SGWR 模型最终的权重矩阵：

$$\gamma = 1 - \alpha, \alpha \in (0, 1] \quad (12)$$

$$W_{GS}(u_0, v_0, x_0) = \alpha * W_G(u_0, v_0) + \gamma * W_S(x_0) \quad (13)$$

在此阶段，引入了参数 α 来控制最终权重矩阵中地理加权矩阵 $W_G(u_0, v_0)$ 和属性加权矩阵 $W_S(x_0)$ 的贡献程度。属性加权矩阵中参数 α 的作用方式与地理加权矩阵中的带宽类似，以 AICc 准则为例，地理加权矩阵需要寻找产生最小 AICc 值的最优带宽，而属性加权矩阵则需要寻找产生最小 AICc 值的最优 α 值。

确定最优 α 值的迭代过程首先将预期范围值 $(0, 1]$ 划分为五个单位间隔的一系列分段，每次迭代时，模型都会选择一个新的 α 值并评估相应的 AICc 值。在迭代过程中，通过降低 AICc 值来逐渐缩小范围。即首先定义一个初始分割函数 $S(\alpha)$ ，它将 α 映射到一系列子范围中：

$$S(\alpha) = \{[1, 0.5], [0.5, 0.1], [0.1, 0.05], \dots\}$$

对于每个子范围 $[a_i, b_i] \in S(\alpha)$ ，通过提取子范围的中点来应用分治算法的步骤：

$$\alpha_i = \frac{a_i + b_i}{2}$$

$$\alpha_{opt} = \arg \min_{\alpha \in (0, 1]} AICc(\alpha) \quad (14)$$

最后, 对于导出的每个 α , 计算相应的 AICc 值, 并且在所有子范围中选择产生最低 AICc 值的最优 α 值。

得到权重矩阵后, 令

$$X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix}$$

$$Y = (y_1, y_2, \dots, y_n)^T, \quad \varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T$$

$$\beta(u_0, v_0, x_0) = [\beta_0(u_0, v_0, x_0), \beta_1(u_0, v_0, x_0), \dots, \beta_p(u_0, v_0, x_0)]^T$$

则目标函数(9)的矩阵形式可以写作

$$\begin{aligned} \min &= (Y - X\beta(u_0, v_0, x_0))^T W_{GS}(u_0, v_0, x_0)(Y - X\beta(u_0, v_0, x_0)) \\ &= Y^T W_{GS}(u_0, v_0, x_0)Y + \beta(u_0, v_0, x_0)^T X^T W_{GS}(u_0, v_0, x_0)X\beta(u_0, v_0, x_0) \\ &\quad - 2\beta(u_0, v_0, x_0)^T X^T W_{GS}(u_0, v_0, x_0)Y \end{aligned} \quad (15)$$

通过对 $\beta(u_0, v_0, x_0)$ 求导可以得到

$$2X^T W_{GS}(u_0, v_0, x_0)X\beta(u_0, v_0, x_0) - 2X^T W_{GS}(u_0, v_0, x_0)Y = 0 \quad (16)$$

最终得到 $\beta(u_0, v_0, x_0)$ 的估计为

$$\hat{\beta}(u_0, v_0, x_0) = [X^T W_{GS}(u_0, v_0, x_0)X]^{-1} X^T W_{GS}(u_0, v_0, x_0)Y \quad (17)$$

将点 (u_0, v_0, x_0) 拓展到点 (u_i, v_i, x_i) ($i = 1, 2, \dots, n$), 从而得到 $\beta(u_i, v_i, x_i)$ 的估计

$$\hat{\beta}(u_i, v_i, x_i) = [X^T W_{GS}(u_i, v_i, x_i)X]^{-1} X^T W_{GS}(u_i, v_i, x_i)Y \quad (18)$$

令 $x_i^T = (1, x_{i1}, x_{i2}, \dots, x_{ip})$ 进而得到 y_i 的拟合值为

$$y_i = x_i^T \hat{\beta}(u_i, v_i, x_i) = x_i^T [X^T W_{GS}(u_i, v_i, x_i)X]^{-1} X^T W_{GS}(u_i, v_i, x_i)Y \quad (19)$$

4. 中国区域人口死亡率影响因素异质性分析

(一) 数据来源与变量描述

本文选取了中国 31 个省份共 337 个地级行政区作为基本研究单位。研究数据主要包括各地级行政区的粗死亡率及其他变量数据和地理空间数据两种类型, 其中地理空间数据主要用于探究地理空间因素对中国地级行政区人口死亡率的影响。本文选取了 2020 年各地级行政区的人口死亡率及其他变量指标, 数据主要来源于人口普查数据、中国统计年鉴及各省的统计年鉴。

最终, 本文选取了人口结构、社会经济、医疗条件方面的 9 项指标来探究中国地级行政区人口死亡率的影响因素, 并对数据进行了标准化处理, 变量定义及描述如表 1 所示。

计算中国地级行政区人口死亡率各个解释变量的方差膨胀因子(VIF), 结果表明, 各变量的 VIF 值均小于 2.5, 即变量之间不存在多重共线性, 可以进行后续分析。

对于 2020 年 337 个中国地级行政区人口死亡率数据的空间分布情况进行分析, 发现中国地级行政区人口死亡率的分布呈现出较为显著的地域差异, 具有空间异质性。研究数据的人口死亡率均值为 5.07‰, 其中, 云南省有较多的地级行政区人口死亡率超过 9‰, 云南省怒江傈僳族自治州的人口死亡率最大, 高

达 13.77‰，广东省深圳市的人口死亡率最小，为 0.98‰。

Table 1. Definition and description of variables
表 1. 变量定义及描述

变量类型	变量	变量描述
人口结构	城镇化率	城镇人口比例
	户口人口比例	户口登记在本地级行政区的人口比例
	人口老龄化程度	65 岁以上人口比重
社会经济	GDP	人均 GDP
	人均住房面积	人均住房建筑面积
	房屋租赁情况	租赁户数占总户数的比例
	零售品消费水平	社会零售品消费占 GDP 的比重
医疗条件	医疗投入	医疗投入占财政支出的比例
	医院床位数	医院卫生院每万人口床位数

(二) 空间自相关分析

空间自相关分析用于衡量研究区域内各位置之间的整体空间相关性及其差异程度，有助于揭示区域之间的空间模式，比如集聚或分散趋势，从而为理解空间分布特征提供依据。莫兰指数(Moran's I)是空间自相关分析的常用指标，其计算公式如下：

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n \sum_{j \neq i}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{S^2 \sum_{i=1}^n \sum_{j \neq i}^n w_{ij}} \quad (20)$$

其中， n 为空间位置的个数， w_{ij} 表示空间位置 i 和 j 的邻近关系，当 i 和 j 邻近时， $w_{ij} = 1$ ；反之， $w_{ij} = 0$ ， x_i 、 x_j 是空间位置 i 和 j 的观察值。

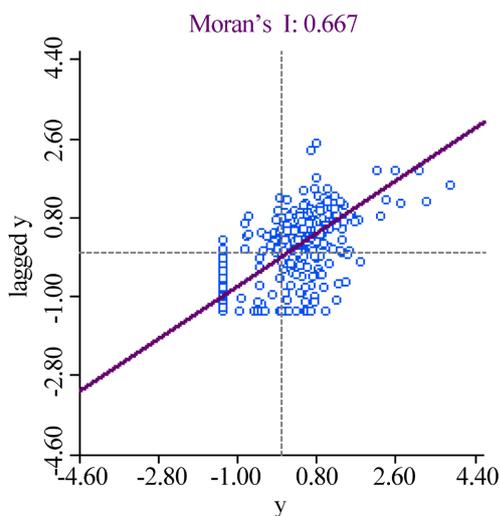


Figure 1. Moran scatter diagram of population mortality in prefecture level administrative regions of China
图 1. 中国地级行政区人口死亡率莫兰散点图

对中国地级行政区的人口死亡率数据进行莫兰检验,探究其空间自相关性,得到散点图如图 1 所示。Moran's I 指数为 0.667,说明中国地级行政区人口死亡率存在显著的空间正自相关性,即相似的人口死亡率在空间上呈现聚集分布模式。

综合上述分析,人口死亡率的波动受地理区域因素的影响较大,具有较为明显的空间异质性与空间依赖性。

(三) 基于 SGWR 模型的异质性分析

本文选择 SGWR 模型对中国地级行政区人口死亡率及其影响因素的空间异质性进行分析,并将回归效果与 GWR 模型进行比较,选用常用的自适应高斯核函数和 AICc 准则建立 GWR 模型和 SGWR 模型。

Table 2. Comparison of fitting effects of different models

表 2. 不同模型拟合效果比较

指标	OLS	GWR	SGWR
R ²	0.354	0.478	0.849
Adj. R ²	0.336	0.439	0.764
AIC	829.187	785.760	562.922
AICc	831.999	789.736	702.395
BIC	-1685.416	878.946	1027.971
RSS	217.751	175.751	50.913
Sigma		0.749	0.485

如表 2 所示,SGWR 模型的 R² 和调整后 R² 均高于 GWR 模型,且 AIC、AICc 以及残差平方和(RSS)均明显低于 GWR 模型。这表明相较于只考虑地理邻近性的 GWR 模型,同时考虑地理邻近性与属性相似性的 SGWR 模型能更有效地捕捉中国地级行政区人口死亡率影响机制的复杂性,整体拟合效果更优。SGWR 模型的系数估计结果如表 3 所示。

Table 3. Statistical characteristics of SGWR model standard regression coefficient estimation results

表 3. SGWR 模型标准回归系数估计结果统计特征

变量	平均值	标准差	最小值	中值	最大值	平均绝对值
截距	-0.101	0.270	-0.534	0.118	1.006	0.234
城镇化率	-0.127	0.271	-1.202	-0.085	0.516	0.215
户口人口比例	-0.038	0.114	-0.384	-0.048	0.495	0.093
人口老龄化程度	0.399	0.210	-0.438	0.433	0.902	0.413
GDP	-0.129	0.189	-0.889	-0.096	0.311	0.168
人均住房面积	-0.042	0.235	-0.788	-0.054	0.634	0.180
房屋租赁情况	-0.247	0.369	-1.461	-0.320	1.610	0.369
零售品消费水平	-0.094	0.102	-0.438	-0.094	0.311	0.113
医疗投入	0.203	0.471	-0.544	0.060	2.546	0.259
医院床位数	0.023	0.361	-0.912	0.009	2.516	0.240

在统计模型中,标准回归系数是指在消除变量单位影响后的系数,它反映了各个自变量对人口死亡

率的影响程度。从标准回归系数的空间分布特征可以看出,各变量对人口死亡率的影响随空间位置的变化而变化。根据标准回归系数的平均绝对值(Mean absolute value, MAV)将 9 个解释变量分为三类:关键变量($MAV \geq 0.3$)包括人口老龄化程度和房屋租赁情况,其影响强度在空间上波动最大,是驱动人口死亡率空间异质性的核心因素;重要变量($0.2 \leq MAV < 0.3$)包括城镇化率、医疗投入和医疗床位数,这些变量也具有明显的空间异质性影响;一般变量($MAV < 0.2$)包括户口人口比例、GDP、人均住房面积和零售品消费水平,其平均影响强度相对较弱,空间波动性也相对较小。接下来,重点分析人口老龄化程度和房屋租赁情况两个关键变量的空间异质性模式及其潜在机制。

基于估计结果,人口老龄化程度的高正相关区域主要集中在中国东部发达地区(如长三角、珠三角核心城市群)以及部分中部省份的核心城市。这些地区虽然经济发达、医疗资源总量丰富,但老年人口基数较大,且高强度的人口集聚导致对医疗、养老等公共资源的竞争加剧,导致老年人口比重增加对死亡率产生显著的正向推动作用。中正相关区域广泛分布于中部大部分地区及部分东北地区。这些区域经济发展水平和医疗资源处于全国中等水平,在应对老年人口增长时,医疗资源、养老服务设施等可能存在相对不足或配置不均衡的问题,使得老年人口比重的增加也会带来死亡率的上升,但强度弱于高正相关区。部分西部地区(如西藏、青海部分地区)和部分东北地区由于优越的自然环境、相对较低的生活压力、以及生活方式等因素,老龄化对死亡率的负面影响可能在一定程度上得到了缓冲,甚至出现负相关(即老年人口比例增加伴随死亡率下降或稳定)。

房屋租赁情况的正相关区域主要集中在西南部分地区(如云南、贵州部分地区)。这些经济发展水平相对滞后、基础设施有待完善的地区,租赁住房居住条件相对较差,身心健康状况易受影响,租赁户比例的增加与人口死亡率的上升呈现关联性。在东部发达地区,租赁市场相对规范成熟,租户有更多选择权,居住环境整体较好。同时,发达的经济带来了更完善的医疗与社会保障。因此,租赁比例的变化对死亡率的影响较小,甚至可能因租户群体相对年轻化等因素而呈现负向影响。其他部分地区则可能由于租赁市场不活跃或其他主导因素更强,使得该变量影响不显著。

(四) SGWR 模型的敏感性分析

为评估 SGWR 模型结果的稳健性,本节从带宽选择和极端值影响两个方面进行敏感性分析。

通过尝试使用交叉验证准则(CV)来选择最优带宽,并与 AICc 准则结果进行对比。结果显示,不同准则选择的最优带宽虽略有差异,但关键变量的空间异质性模式保持相对稳定,模型核心结论未发生本质变化。

Table 4. Fitting effects of different models after removing extreme values

表 4. 去除极端值后不同模型拟合效果

指标	OLS	GWR	SGWR
R ²	0.379	0.507	0.826
Adj. R ²	0.361	0.472	0.742
AIC	811.262	760.339	585.058
AICc	814.079	763.941	694.881
BIC	-1681.449	848.695	1005.858
RSS	208.144	165.277	58.208
Sigma		0.727	0.508

考虑到云南省怒江傈僳族自治州(死亡率 13.77‰)和深圳市(死亡率 0.98‰)为极端值,将其剔除后重新运行 SGWR 模型。结果如表 4 所示,与表 2 相比,SGWR 模型的各项指标呈现小幅波动但整体保持优

越性。具体来看,模型的核心解释力指标 R^2 和调整后 R^2 有所下降,表明极端值的移除轻微降低了模型对因变量变动的解释能力,但 SGWR 仍显著高于 OLS 和 GWR。误差相关指标中,残差平方和(RSS)和残差标准差(Sigma)有所上升,反映极端值剔除后模型残差略有增大,但 SGWR 仍远低于 GWR 和 OLS,说明其预测精度保持相对稳定。

5. 结论及建议

(一) 结论

本文通过构建相似地理加权回归模型(SGWR)来探究中国地级行政区人口死亡率影响因素的空间异质性。根据莫兰指数结果可以得知,中国地级行政区人口死亡率存在显著的空间正自相关性,表明高(低)死亡率地区在空间上呈现聚集分布。

SGWR 模型在拟合人口死亡率数据时,显著优于全局 OLS 模型和仅考虑地理邻近性的 GWR 模型,证明了同时考虑地理邻近性与属性相似性对于捕捉人口死亡率影响机制空间异质性的必要性和优越性。敏感性分析表明,模型结果对带宽的选择和个别极端值具有较好的稳健性,核心结论可靠。

根据 SGWR 模型的系数估计结果可知,人口结构、社会经济、医疗条件三类因素对人口死亡率的影响均存在显著的空间异质性,人口老龄化程度和房屋租赁情况是其中的两个关键变量。人口老龄化程度对死亡率的影响在东部发达地区和中部大部分地区主要表现为正向推动作用,而在西部和东北部分地区影响较弱甚至出现负相关。房屋租赁情况对死亡率的影响在西南欠发达地区表现为显著的正相关影响,在其他部分地区则影响较弱或不显著。

(二) 建议

基于 SGWR 模型揭示的中国区域人口死亡率影响因素空间异质性特征,特别是人口老龄化程度和房屋租赁情况两大关键变量的区域分异规律,本研究提出以下建议:

在人口老龄化程度对死亡率呈现高强度正向推动的东部发达地区,应重点实施提质扩容策略,大力发展养老服务业和专业护理、康复医疗,推动社区嵌入式养老服务,同步推进老年人心理健康干预和慢性病精细化管理;对于老龄化呈中度正向影响的中部及东北地区,应提升县域和基层医疗养老设施与服务能力,推广普惠型养老模式覆盖农村及低收入群体,同时探索区域协作机制共享资源;在西部和东北部分老龄化影响较弱甚至呈负相关的区域,则应充分利用生态优势发展特色康养产业,同时必须筑牢健康基础,持续改善基层医疗条件,确保基本服务可及性。

针对房屋租赁情况与死亡率在西南部分地区的显著正相关,亟需实施租赁住房质量强制准入制度,严格监管居住安全标准,重点改善流动人口聚集区的居住环境;扩大保障性租赁住房供给覆盖低收入租户,并将其纳入本地基本医疗保障体系,加强职业健康防护以降低工作风险。在租赁比例与死亡率呈低相关或负相关的区域,应巩固租赁市场规范化成果,利用数字化手段提升管理效能,持续监测市场变化并防范潜在风险。

此外,还需实施普遍性基础策略以支撑区域行动:持续优化人口结构,综合完善生育支持体系;深化医疗资源供给侧改革,根据区域人口结构与疾病谱变化精准投入资源;促进区域均衡发展,通过产业转移、基础设施互联互通和人才政策引导人口产业合理分布;强化覆盖全民全生命周期的健康教育与疾病预防,提升居民健康素养;建立动态监测评估机制,利用空间统计方法持续跟踪死亡率空间格局变化与政策实施效果,为科学决策提供依据。

基金项目

本文是魏传华主持的国家社科基金项目(21BTJ005)的阶段性成果。

参考文献

- [1] 陈碧珍. 环境视阈下中国省域人口死亡率影响因素实证研究[D]: [硕士学位论文]. 海口: 海南师范大学, 2021.
- [2] Harvey Brenner, M. (1979) Mortality and the National Economy. *The Lancet*, **314**, 568-573. [https://doi.org/10.1016/s0140-6736\(79\)91626-x](https://doi.org/10.1016/s0140-6736(79)91626-x)
- [3] Lee, R.D. and Carter, L.R. (1992) Modeling and Forecasting U.S. Mortality. *Journal of the American Statistical Association*, **87**, 659-671. <https://doi.org/10.1080/01621459.1992.10475265>
- [4] 陶祥兴, 杨峥, 季彦颀. 基于机器学习的 Lee-Carter 模型死亡率预测方法研究[J]. 人口与经济, 2022(6): 47-57.
- [5] Hardarson, T., Gardarsdóttir, M., Gudmundsson, K.T., Thorgeirsson, G., Sigvaldason, H. and Sigfússon, N. (2001) The Relationship between Educational Level and Mortality. The Reykjavík Study. *Journal of Internal Medicine*, **249**, 495-502. <https://doi.org/10.1046/j.1365-2796.2001.00834.x>
- [6] Jusot, (2006) The Shape of the Relationship between Mortality and Income in France. *Annales d'Économie et de Statistique*, **83**, 89-122. <https://doi.org/10.2307/20079164>
- [7] 赵锦辉. 中国人口死亡变化趋势的社会经济因素分析[J]. 西北人口, 1994(4): 49-53.
- [8] 应奎, 李旭东. 中国人口死亡率空间格局演变及其影响因素[J]. 世界地理研究, 2022, 31(2): 440-452.
- [9] Brunson, C., Fotheringham, A.S. and Charlton, M.E. (1996) Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geographical Analysis*, **28**, 281-298. <https://doi.org/10.1111/j.1538-4632.1996.tb00936.x>
- [10] 苏晓丽. 中国五岁以下儿童死亡率影响因素的空间异质性研究[D]: [硕士学位论文]. 西安: 西安科技大学, 2022.
- [11] 赵宇, 魏传华. 一类地理加权 CBD 死亡率预测模型的研究[J]. 数学的实践与认识, 2019, 49(9): 152-156.
- [12] Zhu, A. and Turner, M. (2022) How Is the Third Law of Geography Different? *Annals of GIS*, **28**, 57-67. <https://doi.org/10.1080/19475683.2022.2026467>
- [13] Lessani, M.N. and Li, Z. (2024) SGWR: Similarity and Geographically Weighted Regression. *International Journal of Geographical Information Science*, **38**, 1232-1255. <https://doi.org/10.1080/13658816.2024.2342319>
- [14] Anselin, L. and Li, X. (2020) Tobler's Law in a Multivariate World. *Geographical Analysis*, **52**, 494-510. <https://doi.org/10.1111/gean.12237>