

中国客运量的空间统计分析

曹文彦, 乔 舰

中国矿业大学(北京)理学院统计学系, 北京

收稿日期: 2022年1月18日; 录用日期: 2022年2月5日; 发布日期: 2022年2月18日

摘 要

基于2012~2020年我国客运量以及相应年末常驻人口数量, 本文对我国各省份客运量的空间关联性进行了空间统计分析。通过全域空间自相关分析以及局域空间自相关分析, 明确了我国客运量的重心分布情况: 我国客运量高值集聚区域(热点)集中分布于华东地区和中南地区, 而客运量低值集聚区域(冷点)集中分布于西北地区和东北地区。通过客运量Moran指数的变化, 分析了2020年新冠疫情对于客运量变化产生的影响; 对两年之间的客运量关系以及客运量与年末常驻人口数量进行了双变量空间自相关分析, 得到了相应的结论。

关键词

空间权重矩阵, Moran指数, 双变量自相关

Spatial Statistical Analysis of China Passenger Capacity

Wenyan Cao, Jian Qiao

Department of Statistics, School of Science, China University of Mining & Technology, Beijing

Received: Jan. 18th, 2022; accepted: Feb. 5th, 2022; published: Feb. 18th, 2022

Abstract

Based on the passenger volume of China from 2012 to 2020 and the number of permanent residents at the end of the corresponding year, this paper conducts a spatial statistical analysis on the spatial correlation of passenger volume of each province in China. Through the global spatial autocorrelation analysis and local spatial autocorrelation analysis, the distribution of the center of gravity of passenger volume in China is clarified: the high value of passenger volume agglomeration area (hot spot) is concentrated in east China and central and southern China, while the low value of passenger volume agglomeration area (cold spot) is concentrated in northwest and north-

east China. Through the change of passenger volume Moran index, the impact of COVID-19 on passenger volume change in 2020 is analyzed. The bivariate spatial autocorrelation analysis between passenger volume and resident population at the end of the two years is carried out and corresponding conclusions are obtained.

Keywords

Spatial Weight Matrix, Moran's I, Bivariate Autocorrelation

Copyright © 2022 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

客运量指在一定时期内, 各种运输工具实际运送的旅客数量, 是反映运输业为国民经济和人民生活服务的数量指标, 也是制定和检查运输生产计划、研究运输发展规模和速度的重要指标。本文客运量包括公路客运量、铁路客运量、高铁动车组客运量以及民航客运量。新冠肺炎疫情爆发以来, 人们的生活方式发生了不少变化, 客运量也受到了极大的影响, 尤其是爆发初期的客运量呈现断崖式下降, 同时客运量也呈现出一定的区域差异、省际差异, 与各省常住人口也存在一定的关联性。

2. 研究对象与数据来源

本文以中国一带一路网(<https://www.yidaiyilu.gov.cn/jcsjpc.htm>)官网公布的 2012~2020 年 31 个省级行政区(港澳台除外)客运量为研究对象, 同时基于 2018 年和 2020 年的各省年末常住人口作相关性分析[1]。相应数据显示: 江苏、四川、湖南、河南、广东为 9 年中客运量始终处于前 5 名的省份, 排名后 3 位的为宁夏、青海、西藏。2012~2020 年全国客运量平均值分别为 121680.58 万人、67341.77 万人、70006.29 万人、61279.13 万人、59722.58 万人、57856.90 万人、55891.81 万人、54659.39 万人、29831.23 万人, 呈现年客运量逐步减少的趋势, 尤其 2013 年和 2020 年呈现断崖式下降。

3. 研究方法

3.1. 空间权重矩阵

空间权重矩阵是空间数据分析的核心要素, 是进行空间统计分析的基础。空间权重矩阵主要分为三大类: 基于地理邻近性的空间权重矩阵、基于空间距离的空间权重矩阵、基于社会经济结构的空间权重矩阵。本文基于地理 Rook 邻近性构建空间权重矩阵。该矩阵适合于测量地理空间效应的影响, 即适合本文中由于地理空间的原因, 使各省客运量出现省际差异; 同时适用于分析空间溢出效应, 分析一个省份对其邻近省份以及邻近省份的邻近省份客运量的影响。

若用 l_{ij} 来表示省份 i 与省份 j 的共有边界长度, 地理 Rook 空间权重矩阵定义为:

$$\omega_{ij} = \begin{cases} 1, & l_{ij} > 0 \\ 0, & l_{ij} = 0 \end{cases}, \quad i, j = 1, 2, \dots, 31 \quad (1)$$

当共有长度大于 0 时, 认为相应省份相邻; 若共有长度等于 0, 认为相应省份不相邻。

3.2. 全域空间自相关

全域空间自相关反映客运量在全国范围内的空间分布模式。本文通过 Moran 指数从整体上研究我国各省份客运量的集聚分布情况, 检验 31 个省份的客运量是否存在空间自相关。Moran 指数的取值为[-1,1]。若 Moran 指数的取值为(0,1], 说明各省客运量存在聚集关系, 且越大, 该省的客运量与邻近省份客运量的相关性越明显[2]。

Moran 指数定义如下:

$$Moran's\ I = \frac{N \sum_{i=1}^N \sum_{j=1}^N \omega_{ij} (y_i - \bar{y})(y_j - \bar{y})}{S_0 \sum_{i=1}^N (y_i - \bar{y})^2} \quad (2)$$

其中, N 表示省份的数量; y_i 表示第 i 个省份的客运量; \bar{y} 表示变量 y_i 的均值; ω_{ij} 表示 3.1 中基于 Rook 邻近性的空间权重矩阵; 可基于相应 Moran 指数的标准化变量 z 实现原假设为各省份客运量随机分布的显著性假设检验。

3.3. 局域空间自相关

局域空间自相关反映在研究区域中某省份与邻近省份关于客运量的相关程度, 局域 Moran 指数基于相邻省份客运量与全国客运量均值进行比较而确定[1]。本文采用 LISA 统计量研究一个省份的客运量与其相邻省份客运量之间的差异程度以及相应的显著性。

局域 Moran 指数定义如下:

$$Local\ Moran's\ I_i = \frac{y_i - \bar{y}}{S_i^2} \sum_{j=1, j \neq i}^N \omega_{ij} (y_j - \bar{y}) \quad (3)$$

其中, N 表示省份的数量; y_i 表示第 i 个省份的客运量; \bar{y} 表示变量 y_i 的均值; ω_{ij} 表示 3.1 中基于 Rook 邻近性的空间权重矩阵。

4. 客运量空间分析

本文分析客运量的空间自相关性, 采用基于地理邻近性的空间权重矩阵, 首先生成一阶基于 Rook 邻近性的空间权重矩阵, 对客运量进行空间自相关分析, 主要通过 Moran 指数以及 Moran 散点图来实现。

4.1. 2018 年客运量全域自相关分析

首先对我国 2018 年客运量进行置换次数为 999 次的原假设为全局自相关系数为零的显著性检验, 结果如表 1 所示。由表 1 可知, Z 值为 1.8972, P 值为 0.0390, 因此在 5% 的显著性水平下拒绝了原假设: 我国客运量呈现随机分布。同时得到 Moran 指数 = 0.1853, 说明我国 31 个省份的客运量在空间分布上具有显著的正向空间自相关, 即存在空间依赖性。说明省域的客运量在空间分布上并不是随机的, 而是表现出某些省份的客运量在空间上趋于集聚的现象, 即客运量较高的省份趋于和客运量较高的省份靠拢的空间分布状态。

Table 1. Global autocorrelation significance test of national passenger capacity in 2018

表 1. 2018 年全国客运量全域自相关显著性检验

Moran 指数	Z 值	P 值
0.1853	1.8972	0.0390

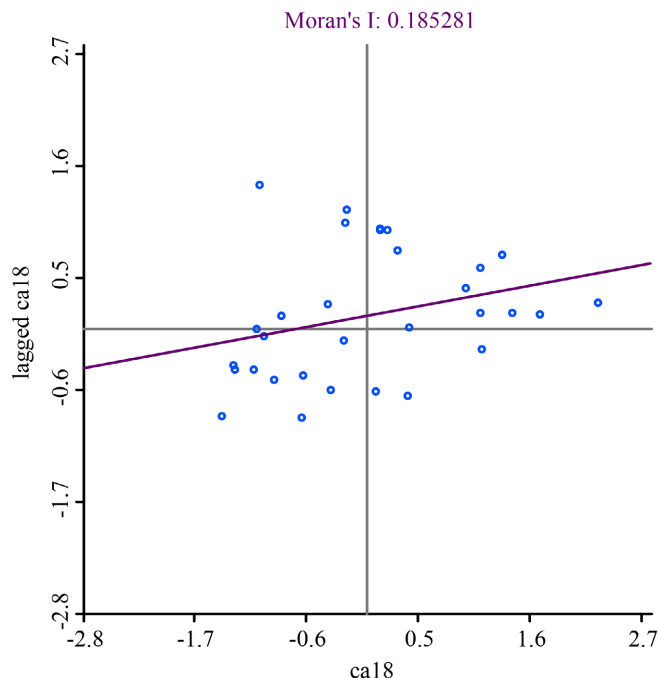


Figure 1. Moran scatter chart of the passenger capacity of China's 31 provinces in 2018

图 1. 2018 年中国 31 个省份客运量 Moran 散点图

全局 Moran 指数散点图如图 1 所示。Moran 散点图中，斜线的斜率为 0.1853，直角坐标的 X 轴为客运量的标准化取值(简称为 pas_cap)，Y 轴为客运量变量的空间滞后项的标准化取值(简称为 $W \cdot pas_cap$)，同时 Moran 散点图中的点对应相应的省份，样本点距离原点的远近代表了聚集显著性的高低。一、三象限表示正的空间自相关，相似值集聚，二、四象限不存在空间自相关。即第一象限的 12 个省份：陕西、山东、河南、安徽、湖北、湖南、江西、贵州、重庆、广东、江苏、浙江存在正的空间自相关，该 12 个省份为高客运量，同时邻近省份也是高客运量；第三象限的 10 个省份：黑龙江、内蒙古、西藏、吉林、甘肃、天津、宁夏、青海、新疆、河北，也存在正的空间自相关，但第三象限表示本身是低客运量，同时也被低客运量的省份所包围。综上所述，2018 年的高客运量省份集中分布于华东地区和中南地区，而低客运量省份集中分布于西北地区 and 东北地区。位于二、四象限的 9 个省份(山西、云南、福建、广西、上海、海南、北京、四川、辽宁)不存在空间自相关。例如，北京处于第四象限，为高客运量，但其邻近省份天津和河北均为低客运量；对于第二象限中的离群点——上海，该省份为低客运量，但其邻近省份是高客运量的江苏和浙江。

4.2. 2018 年客运量局域自相关分析

Moran 散点图给出了各省份所属的象限，但不能判断各个省份的局域相关类型及其聚集区域是否在统计意义上显著，因此需要使用 LISA 集聚图来进一步分析。

接下来对 2018 年 31 个省份的客运量进行局域自相关分析。局域自相关分析对显著性检验比较敏感，检验结果非常依赖于置换的次数和显著性水平，因此在结果分析之前需要做敏感性分析，即进行多次置换检验。

对我国 2018 年客运量进行置换次数为 999 次的原假设为局域自相关系数为零的显著性检验，结果如表 2 所示。

Table 2. Local autocorrelation significance test of national passenger capacity in 2018
表 2. 2018 年全国客运量局域自相关显著性检验

指标	P = 0.05	P = 0.01	P = 0.01
Moran 指数	新疆、内蒙古、甘肃、湖北、湖南、 广西、重庆、福建、上海	安徽、江西	无

同时得到 2018 年各省份客运量聚类地图, 如图 2 所示:



Figure 2. LISA cluster map of passenger capacity by province in 2018

图 2. 2018 年各省份客运量聚类地图

表 2 为局域自相关检验统计上显著的省份, 共有 11 个, 其余省份在统计上不显著。该 11 个省份在地图中表现为聚类的核心区域, 如图 2 所示。在聚类地图中, 较黑的区域包含 5 个省份, 即重庆、湖北、安徽、湖南、江西, 表示显著高 - 高聚类; 灰色区域有 3 个省份, 即上海、福建、广西, 表示显著低 - 高聚类; 黑色区域有 3 个省份, 即新疆、甘肃、内蒙古, 表示显著低 - 低聚类。由图 1 可知, 高 - 高聚类有 12 个省份, 而 LISA 集聚图中只有 5 个省份, 说明重庆、湖北、安徽、湖南、江西这 5 个省份的客运量在统计意义上更为显著, 其余 7 个省份虽然存在聚类现象, 但不够显著。可以明显地看出, 高客运量省份集中分布于长江中游地区, 低客运量省份聚集在西北地区, 显然这与经济发展程度和人口分布有直接关系。

4.3. 2020 年客运量全域自相关分析

众所周知, 2019 年底、2020 年初, 新冠疫情大面积爆发, 对我国的经济、对人们的生活等都造成了严重的影响[3]。同样地, 对我国客运量也造成了一定的冲击, 2020 年 31 个省份的客运量均值仅占 2018 年 31 个省份客运量均值的 53.37%, 出现了断崖式下降。

下面对 2020 年我国 31 个省份的客运量进行空间统计分析。

Table 3. Global autocorrelation significance test of national passenger capacity in 2020
表 3. 2020 年全国客运量全域自相关显著性检验

Moran 指数	Z 值	P 值
0.1658	1.7567	0.0470

对我国 2020 年客运量进行全域空间自相关显著性检验, 得到表 3。结果显示, 我国 31 个省份的客运量在空间分布上具有显著的正向空间自相关, 即客运量较高的省份趋于和客运量较高的省份靠拢的空间分布状态。对比表 1 和表 3, 2020 年的客运量 Moran 指数较 2018 年明显地减少。疫情爆发后, 人们纷纷居家隔离、居家办公, 很大程度上减少出行, 生活方式被改变的同时, 我国客运量也大量降低, 最终导致各省份客运量之间的相关程度减少。

接下来给出 2020 年各省份客运量 Moran 指数散点图。图 3 中各坐标系以及斜线斜率等与图 1 中所代表含义相同。

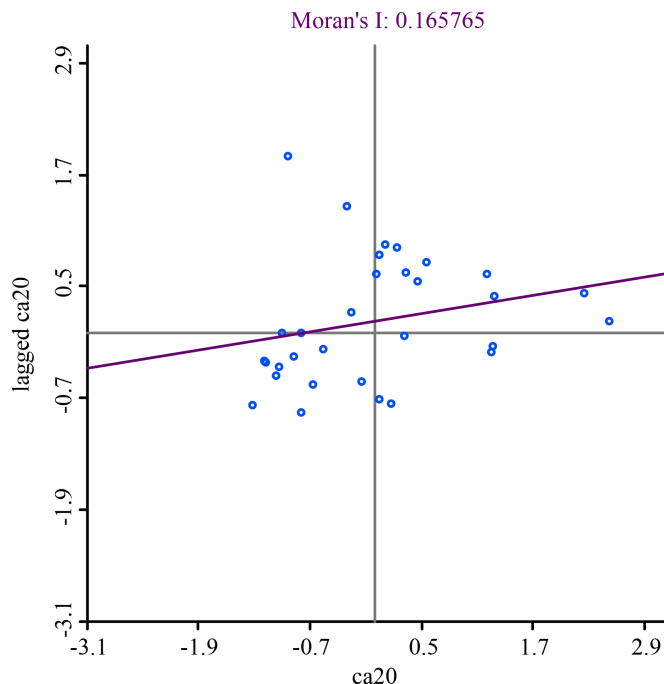


Figure 3. Moran scatter chart of the passenger capacity of China's 31 provinces in 2020

图 3. 2020 年中国 31 个省份客运量 Moran 散点图

在图 3 中, 较 2018 年中国 31 个省份客运量 Moran 散点图, 广西省从第二象限变为了第一象限, 由低 - 高聚类变为了高 - 高聚类; 陕西省和河南省由第一象限变为了第四象限, 高 - 高聚类变为高 - 低聚类。云南省始终位于第三象限, 说明该省份的客运量始终处于相对较低的水平, 然而四川省始终处于第四象限, 因此应充分发挥云南省旅游业, 由邻省四川发挥带动作用, 增加客运量, 从而促进该区域经济发展。虽然个别省份的象限发生变化, 但在受到疫情影响后, 2020 年的高客运量省份仍集中分布于华东地区和中南地区, 低客运量省份仍集中分布于西北地区 and 东北地区。

4.4. 2020 年客运量局域自相关分析

接下来对 2020 年中国 31 个省份的客运量进行局域空间自相关分析。

对我国 2020 年客运量进行置换次数为 999 次的原假设为局域自相关系数为零的显著性检验, 结果如表 4 所示。

同时得到 2020 年各省份客运量聚类地图, 如图 4 所示。

Table 4. Local autocorrelation significance test of national passenger capacity in 2020**表 4.** 2020 年全国客运量局域自相关显著性检验

指标	P = 0.05	P = 0.01	P = 0.01
Moran 指数	新疆、内蒙古、湖南、广西、 山东、福建、江西	安徽、上海	无

**Figure 4.** LISA cluster map of passenger capacity by province in 2020**图 4.** 2020 年各省份客运量聚类地图

表 4 为局域自相关检验统计上显著的省份, 共有 9 个省份, 其余省份在统计上不显著。甘肃、湖北、重庆相较于 2018 年不再显著。2020 年湖北、重庆、甘肃受疫情影响较为严重, 且湖北省武汉市出现了封城现象, 因此这三个省份的客运量受到较为明显的影响。在聚类地图中, 广西省在检验统计上显著且变为高-高聚类, 与全域自相关分析中一致, 由两年的客运量数值比较可知, 广西省客运量减少的比例最小, 因此产生了上述变化; 上海、福建始终为低-高聚类。

4.5. Moran 指数及客运量变化分析

由上述分析可知, 2020 年较 2018 年 31 个省份的客运量 Moran 指数发生了变化, 表 5 为 2012~2020 年相应全国客运量 Moran 指数汇总。2013~2020 年的 P 值均小于 0.1, 则说明在 10% 的显著性水平下均有效。而 2012 年的 P 值大于 0.1, 接受原假设: 2012 年我国客运量呈现随机分布。

Table 5. Moran's I from 2012 to 2020**表 5.** 2012~2020 年 Moran 指数

year	Moran 指数	Pvalue	Zscore
2012	0.0719	0.158	0.9687
2013	0.1614	0.045	1.7676
2014	0.1800	0.046	1.8030
2015	0.1418	0.062	1.5822
2016	0.1627	0.054	1.723
2017	0.1803	0.042	1.8467
2018	0.1853	0.031	1.9503
2019	0.1666	0.053	1.6971
2020	0.1658	0.052	1.7193

图5为2013~2020年Moran指数变化折线图。可以看出,有三个明显的转折点分别发生在2014年、2015年和2018年, Moran指数不断发生变化,同时2020年的Moran指数较2019年再次减小,但是依然呈现空间自相关。

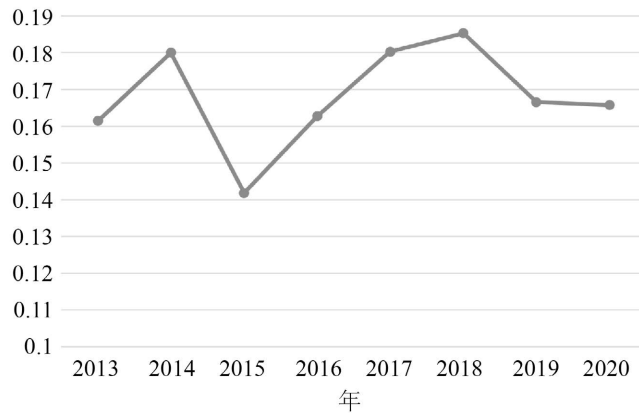


Figure 5. A line chart of the Moran's I from 2013 to 2020
图5. 2013~2020年Moran指数折线图

同时分析2012~2020年全国各省份的客运量变化,如图6所示。2012~2020年的客运量依次逐年下降,尤其是2013年和2020年,客运量较前一年均大幅度减少。查阅资料可知,2013年我国的公路客运量统计口径发生了变化,以车辆属性为判断依据,仅关注营业性运输的内容,也不再将公路里程超过总行程一半的趟次纳入公路运输量统计范围,导致2013年客运量暴跌[4]。2019年底,疫情爆发,对交通产生了很大的冲击,表现为客运量大幅减少。《联合早报》报道:国际航空运输协会(IATA)分析报告披露,疫情对航空业的冲击已经延长至中期,其中,预计长途和国际航班受创最重。它估计2021年全球客运需求将比2019年少24%,客运量要到2023年才有望超过2019年的水平[5]。疫情对公路客运量、铁路客运量等的影响也是如此,大大限制了人们的出行以及交通工具的乘坐。

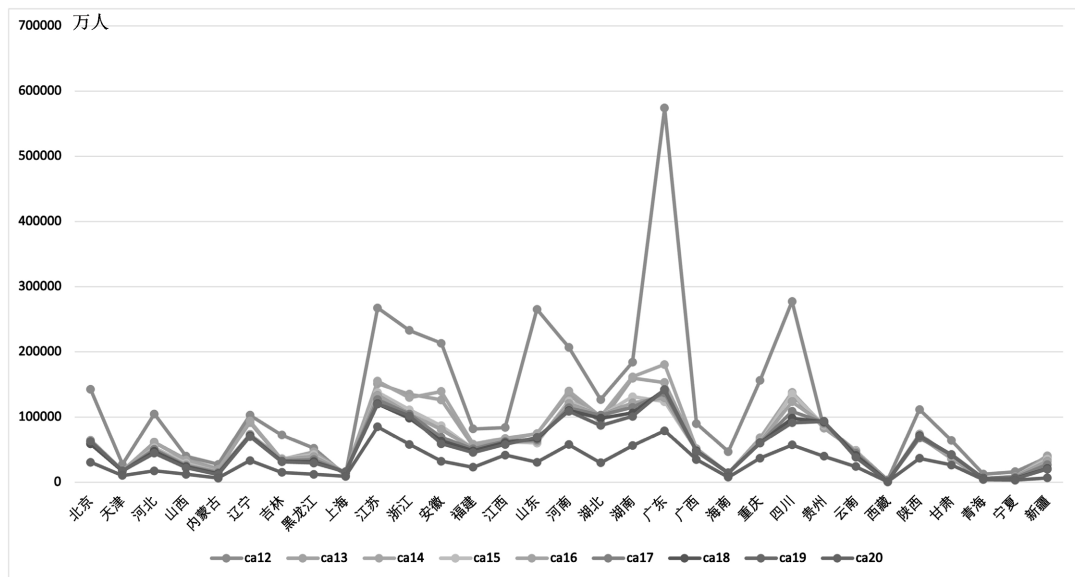


Figure 6. Line chart of passenger capacity from 2012 to 2020
图6. 2012~2020年客运量折线图

4.6. 客运量中心及走向分析

图 7 为 2015~2020 年我国客运量的四分位图, 颜色由浅及深表示客运量由少增多。由图 7 可以看出客运量较高的省份集中在华东地区和中南地区, 同时由图 6 也可以看出, 颜色极深的省份的客运量在 2015~2020 年处于较高的水平, 并且这些省份在 Moran 散点图中处于第一象限, 属于高-高聚类, 因此该区域为客运量的重心区域。客运量低的省份集中分布于东北地区 and 西北地区。综上可知, 我国各省份的客运量具有显著的空间异质性。

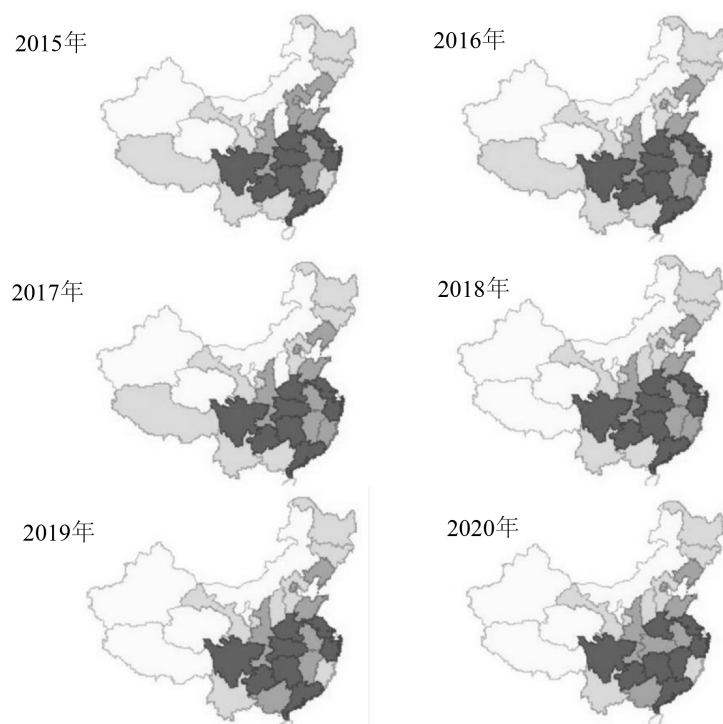


Figure 7. Passenger capacity of China's 31 provinces from 2012 to 2020
图 7. 2012~2020 年中国 31 个省份客运量层次图

5. 双变量空间自相关分析

5.1. 客运量与年末常住人口

接下来研究各省客运量与该省邻近省份的年末常住人口加权均值的空间交互关系。对 2018 年和 2020 年分别进行双变量分析。结果如表 6 所示:

Table 6. Bivariate correlation analysis

表 6. 双变量自相关分析

年份	双变量 Moran 指数	显著性检验		LISA 聚类地图	
2018 年	0.2022	P 值	0.0230	高 - 高	山东河南安徽江西
		Z 值	2.1888	低 - 低	新疆甘肃
2020 年	0.1923	P 值	0.0400	高 - 高	山东安徽江西
		Z 值	1.9602	低 - 低	新疆甘肃

由表 6 可知, 2018 年双变量 Moran 指数的值为 0.2022, P 值为 0.0230; 2020 年双变量 Moran 指数为 0.1923, P 值为 0.0400, 在统计上显著为正, 因此二者存在正向空间相关关系。因此, 各省的客运量与邻近省份的年末常住人口加权均值存在一定的正向空间相关关系。

分析可知, 双变量 Moran 散点图中各象限所对应的省份与客运量全域空间自相关 Moran 散点图中所对应的省份基本一致, 依然是在华东和中南地区呈现高客运量 - 高人口分布, 在东北和西北地区呈现低客运量 - 低人口分布。得益于四川省旅游业蓬勃发展, 尽管其周围省份青海、甘肃、新疆的常住人口较少, 但是依然拥有较高的客运量, 西北地区人口稀少, 客运量也始终处于较少程度, 因此应大力发展旅游业, 发挥其自身优势, 从而带动客运量的增加, 进一步带动经济的发展。

5.2. 两年客运量之间的相关分析

接下来通过双变量的 Moran 指数分析两年客运量之间的关系。

首先, 2018 年各省客运量(简称为 pas_cap2018)作为第一变量(X), 2020 年各省客运量(简称为 pas_cap2020)作为第二变量(Y), 结果如图 8 所示:

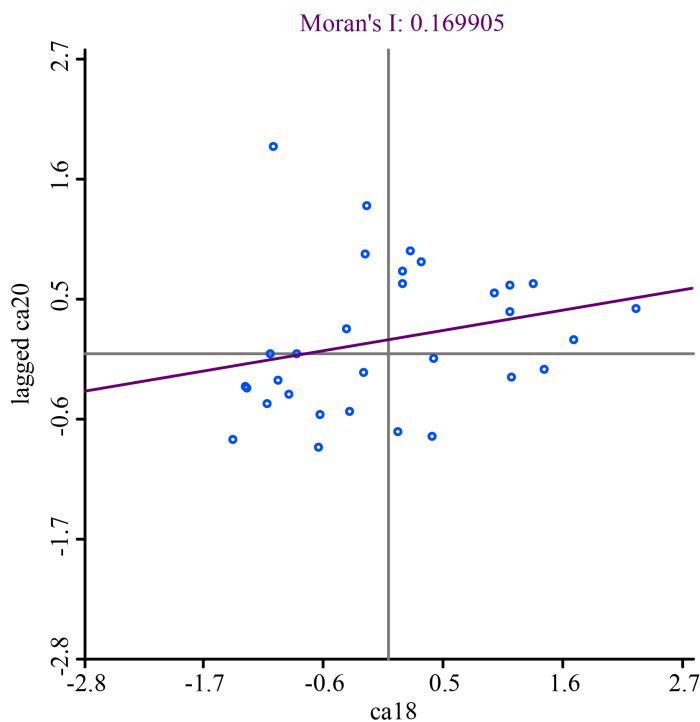


Figure 8. Bivariate Moran scatter chart

图 8. 双变量 Moran 散点图

图 8 描述的是一个省份 2018 年的客运量和周围邻居省份 2020 年加权平均客运量的时空关系。由时间的单向影响关系可知, 这个 Moran 散点图表示一个省份 2018 年的客运量对它的邻居省份 2020 年加权平均客运量的空间交互影响, 即外向溢出, 由过去的中心地区溢出到现在周围的地区。由图可知, Moran 指数为 0.1699, 因此, 一个省份 2018 年的客运量对它的邻居省份 2020 年加权平均客运量会产生正向的影响。

其次, pas_cap2020 作为第一变量(X), pas_cap2018 作为第二变量(Y), 得到图 9:

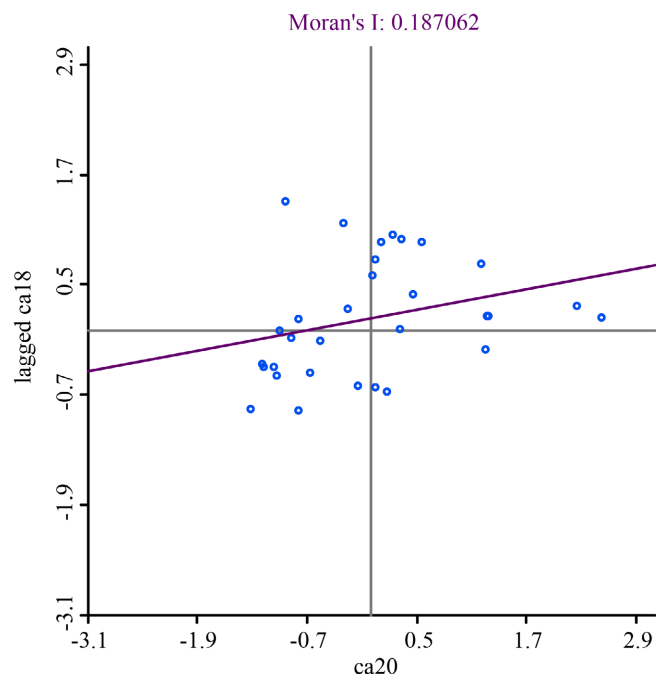


Figure 9. Bivariate Moran scatter chart
图 9. 双变量 Moran 散点图

如图 9 描述的是一个省份 2020 年的客运量和周围邻居省份 2018 年加权平均客运量的时空关系, 即一个省份的 2020 年客运量受到邻居省份 2018 年加权平均客运量的空间交互影响, 这可以看作是内向溢出, 由过去的邻居地区向现在的核心地区溢出。由图可知, Moran 指数为 0.1871, 则 2018 年某省份的邻居省份加权平均客运量会对该省份 2020 年的客运量产生正向影响。

由上述分析可知, 某省前一年的客运量会与该省份的邻居省份的未来客运量产生正向的相关关系; 同时某省的邻居省份前一年的加权平均客运量会对该省份未来年份的客运量产生正向相关影响。

6. 结论

本文利用 2012~2020 年 9 年的客运量以及年末常住人口, 对我国各省份客运量的空间关联性进行空间统计分析, 得到如下结论:

- 1) 我国的高客运量省份集中分布于华东地区和中南地区, 低客运量省份集中分布于西北地区 and 东北地区, 与人口分布情况相一致。
- 2) 得益于旅游业的蓬勃发展, 四川省出现了高客运量 - 低人口的分佈情况, 因此四川省应该充分带动其周围省份云南、青海、甘肃、新疆大力发展旅游业, 发挥自身优势, 推动经济发展。
- 3) 我国各省份 2012~2020 年 9 年的客运量逐年下降。由于疫情的爆发, 人们的出行被强制减少, 2020 年的客运量出现断崖式下降, 由此带来的经济效益也受到了严重影响。因此应加大力度进行疫情防控, 适当调整乘客出行方式, 促使经济稳步发展。
- 4) 各省前一年的客运量会与未来该省份邻居的客运量产生正向的相关关系。因此, 客运量的增加或减少不是一蹴而就的, 但当遇到疫情这种罕见的突发情况时, 也会发生断崖式变化。

参考文献

- [1] 乔舰, 谭佳宁. 中国邮政快递业收入区域差异研究[J]. 统计学与应用, 2021, 10(2): 345-353.

<https://doi.org/10.12677/SA.2021.102034>

- [2] 庞新怡, 富丽, 徐佳, 蒋青艳. 区域性金融风险向宏观系统性金融风险的空间传导路径——以江苏省为例[J]. 中国商论, 2019(9): 51-53.
- [3] 廖沈美慧, 陈春娇, 张知青, 刘伟. 新冠肺炎疫情对地铁客运量的影响分析[J]. 城市轨道交通, 2021(1): 53-56.
- [4] 贲莉莉. 营业性公路运输量统计口径变迁及测算方法探讨[J]. 江苏交通科技, 2015(3): 23-25.
- [5] 联合早报. 国际航协: 疫情影响料持续多年 2023 年前难恢复至去年客运量[EB/OL]. <http://sg.mofcom.gov.cn/article/dtxx/202005/20200502965307.shtml>, 2020-05-16.