

Analysis and Prediction of Passenger Flow Differences in Shanghai 3A Scenic Spots and Above Based on ARIMA Model

Shiyao Song, Yuchun Zhou

Shanghai University of International Business and Economics, Shanghai
Email: 479076815@qq.com

Received: Jun. 5th, 2019; accepted: Jun. 19th, 2019; published: Jun. 26th, 2019

Abstract

In recent years, with the rapid development of urban tourism in Shanghai, the passenger flow of major scenic spots is increasing day by day. In order to further understand the changing characteristics of the passenger flow of Shanghai scenic spots in different periods, such as holidays and working days, this program counts 35 scenic spots of 3A level and above in Shanghai from March 15, 2018 to October 15, 2018 through the official Wechat public number of Shanghai Tourism Bureau, "Happy Tour Shanghai". The daily passenger flow is 10:00 and 15:00. On the basis, through descriptive statistical analysis, it is concluded that the daily passenger flow at 15:00 is generally higher than that at 10:00 in Shanghai tourist attraction. K-means cluster shows that several popular scenic spots occupy most of the daily passenger flow in Shanghai, while most of the 3A and 4A scenic spots are still at a low level. Through time series analysis of passenger flow in different scenic spots, ARIMA model is established. It is concluded that there are different fluctuations of passenger flow in scenic spots 3A and above in Shanghai, and there are mutual influences in time. Through descriptive statistical analysis, K-means clustering analysis and time series modeling, graphics and models coexists; this project analyzed the distribution of passenger flow intuitively and scientifically, and predicted the future trend of tourist flow in scenic spots, hoping to provide help for the development of Shanghai tourism.

Keywords

Passenger Flow in Scenic Spots, Shanghai Tourist Scenic Spot, K-Means Cluster Analysis, Time Series Analysis, ARIMA Model

基于ARIMA模型的上海3A级及以上景区客流量差异分析及其预测研究

宋诗瑶, 周玉春

上海对外经贸大学, 上海
Email: 479076815@qq.com

收稿日期: 2019年6月5日; 录用日期: 2019年6月19日; 发布日期: 2019年6月26日

摘要

近年来, 上海的都市旅游发展迅速, 各大景区客流量也出现日益增多的情况, 为了进一步了解上海景区在节假日和工作日等不同时段的客流变化特征, 本项目通过上海市旅游局官方微信公众号“乐游上海”统计了2018年3月15日至2018年10月15日上海市35个3A级及以上景区每日10:00和15:00客流量, 在此基础上, 通过描述统计分析得出上海旅游景区每日15:00客流量普遍高于10:00客流量。通过K-means聚类分析发现几个热门景点占据了上海每日大部分客流量, 而大部分3A及4A景区客流量仍处于较低水平。通过对不同类型的景区客流量进行时间序列分析, 建立ARIMA模型, 得出上海市3A及以上景区客流量存在不同的波动性, 并且在时间上存在相互影响。本项目通过描述性统计分析, K-means聚类分析及时间序列建模的方法, 图形与模型并存, 直观科学地分析了客流量的分布情况并预测了景区客流量的未来趋势, 希望能对上海旅游业发展提供帮助。

关键词

景区客流量, 上海旅游景区, K-Means聚类分析, 时间序列分析, ARIMA模型

Copyright © 2019 by author(s) and Hans Publishers Inc.

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

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



Open Access

1. 引言

1.1. 背景及目的

1978年国家为改革开放大局所需提出“大力发展旅游事业”, 由此, 我国旅游业开始萌芽。到1998年12月, 中央经济工作会议把旅游业明确为“国民经济新的增长点”。随后, “假日制度”推出, 大众旅游风生水起, 旅游市场繁荣兴旺。而2017年, “大力发展全域旅游”成为中央经济工作会议对旅游业定位的重要肯定。

携程发布的《2018上海旅游大数据报告》显示, 2017年上海共接待国际旅游入境者873.01万人次, 国内旅游者31845.27万人次, 2018年同比增长55%。文化和旅游部发布的2018年旅游市场基本情况显示, 2018年全年, 国内旅游人数55.39亿人次, 比上年同期增长10.8%。

在旅游业繁荣发展的情况下, 部分旅游景区在法定节假日也出现了客流量激增的情况, 大大降低了游客的游玩体验。为了旅游业更长远的发展, 同时符合可持续发展战略, 对各景区客流量的研究和预测是极其重要的。

上海作为经济领先的大都市, 旅游业也发展迅速, 但同时也不可避免的出现了景区客流量日益增多的情况。在寒暑假、法定节假日, 上海部分景点均出现客流量趋于饱和的现象。由此, 本文基于上海市3A及3A级以上景区的客流量数据, 分别比较不同日期和不同时间段上海旅游景点客流量差异; 分析法定节假日、寒暑假及上海旅游节对上海景点总客流量的影响; 对各景点依据客流量进行聚类分析, 并建

立时间序列模型预测未来的客流量, 为上海旅游业的可持续发展提出有效的优化方案。

1.2. 研究方法

本文运用的主要研究方法是:

- 1) 观察法: 通过上海市旅游局官方微信公众号——乐游上海观测各景点每日客流量数据并到各景点进行实地考察。
- 2) 文献研究法: 通过查找与景区客流量、时间序列等相关的文献, 获得相关的研究方法及研究经验。
- 3) 描述性统计分析: 绘制节假日和非节假日及不同时间段的客流量折线图, 直观形象的对比分析节假日及时间段对客流量的影响。
- 4) K-means 聚类分析: 对 35 个不同的旅游景点的客流量进行分类, 分析客流量相似的部分景点。
- 5) 时间序列建模: 建立时间序列 ARIMA 模型, 分析景点客流量在时间上的变化趋势。

1.3. 研究现状

近年来, 随着国内旅游业不断发展, 如何预测景区客流量成为重要研究问题。目前, 国内在客流量预测领域的研究主要集中在城市公共交通而对景区客流量预测的研究较少。对上海旅游客流量的研究调查主要有旅游承载力研究[1] (2007, 汪宇明等)以访谈及问卷调查的方式进行实证分析[2] (舒锡慧等, 2011), 上海市旅游景区点客流时空分布特征及成因分析[3] (刘春济等, 2006 年)和景点管理可容纳量的实际可操作性, 并强调利用政策和制度的强制性来确保对旅游景点的管理工作能顺利进行[4] (黄筱焯, 2005 年)。近年来, 对旅游客流量的分析已经加入了统计分析方法, 主要有基于 ARIMA 模型的旅游业实证研究[5] (陈珊等, 2017), 对年度旅游数据建立 ARIMA 模型进行预测分析[6] (刘燕玉等, 2016 年), 运用不用时间维度的景点人数进行时间序列分析[7] (赵玥, 2018 年)。本文基于以上研究, 对上海市的旅游日度数据进行时间序列分析, 建立 ARIMA 模型并进行预测, 试图用模型展示上海市的每日旅游量的变化趋势。

1.4. 数据来源与处理

通过上海市旅游局官方微信公众号——乐游上海统计了 2018 年 3 月 15 日至 2018 年 10 月 15 日上海市 87 个 3A 级及以上景区每日 10:00 和 15:00 客流量。由于部分园区存在季节性闭园、部分时间段数据缺失的情况, 最终筛选了如表 1 所示的 35 个数据完整的景区进行研究分析。在此基础上, 将所有数据分为节假日和非假日, 并将每一天 35 个景区的数据求和, 从宏观及微观数据上分析上海市 3A 及 3A 级以上景区的客流量。

2. 景区客流量描述性统计分析

2.1. 不同时间段的景区客流量差异分析

在以下的描述性统计分析中, 将已筛选出的 35 个旅游景点客流量总和作研究对象, 建立日客流量折线图。

通过图 1 和图 2 两张日客流量折线图可以较为直观地观察到旅游景区每日 10:00 和 15:00 客流量存在显著性差异且 15:00 客流量普遍高于 10:00 客流量。

对节假日 10:00 和 15:00 客流量进行方差齐性检验。如图 3 所示, 其中 A 代表 10:00, B 代表 15:00。两样本的方差齐性 F 检验的原假设为“ A 和 B 样本的方差相等”, 对应的 P 值 < 0.0001 , 则拒绝原假设, 认为两组观测的方差差异显著, 可以证明 10:00 和 15:00 景区客流量的确存在显著性差异。

相对于早上, 游客更多的会选择下午的时间段到景区游玩, 另外, 导致下午景区客流量高于上午的原因可能是游客在景区停留的时间较长, 所以到下午, 景区的客流量会越来越多。

Table 1. Overview of passenger flow in major scenic areas
表 1. 各大主要景区客流量概览图

| | 2018.3.15 | | 2018.3.16 | | ... | 2018.10.15 | |
|--------------------|-----------|--------|-----------|--------|-----|------------|--------|
| 上海国际旅游度假区 | 4446 | 18,464 | 19,543 | 22,544 | ... | 10,845 | 22,791 |
| 东方明珠广播电视塔 | 123 | 1023 | 888 | 1215 | ... | 266 | 999 |
| 上海野生动物园 | 128 | 476 | 2591 | 1640 | ... | 837 | 3954 |
| 上海科技馆 | 120 | 788 | 2012 | 1076 | ... | 536 | 239 |
| 上海自然博物馆 | 63 | 467 | 1207 | 432 | ... | 263 | 834 |
| 上海博物馆 | 174 | 678 | 616 | 918 | ... | 323 | 1171 |
| 上海佘山国家森林公园 西佘公园 | 129 | 67 | 3 | 115 | ... | 291 | 713 |
| 上海佘山国家森林公园 东佘公园 | 25 | 86 | 58 | 106 | ... | 14 | 0 |
| 上海豫园 | 377 | 678 | 413 | 845 | ... | 700 | 613 |
| 上海城市规划馆 | 56 | 497 | 61 | 338 | ... | 110 | 178 |
| 上海动物园 | 330 | 636 | 512 | 574 | ... | 683 | 1549 |
| 上海世纪公园 | 340 | 876 | 551 | 1058 | ... | 398 | 427 |
| 上海大观园 | 122 | 89 | 168 | 54 | ... | 201 | 4 |
| 朱家角古镇 | 531 | 980 | 1416 | 1010 | ... | 840 | 1127 |
| 上海市青少年校外活动营地——东方绿舟 | 32 | 299 | 1090 | 1706 | ... | 82 | 119 |
| 上海海洋水族馆 | 496 | 522 | 541 | 451 | ... | 657 | 468 |
| 上海鲜花港 | 42 | 17 | 32 | 32 | ... | 6 | 98 |
| 上海月湖雕塑公园 | 57 | 34 | 79 | 35 | ... | 35 | 2144 |
| 上海嘉定州桥 | 2872 | 1520 | 3948 | 2748 | ... | 1220 | 1140 |
| 上海长风公园 | 144 | 465 | 1144 | 1147 | ... | 1462 | 1758 |
| 前卫生态村 | 12 | 5 | 70 | 25 | ... | 14 | 61 |
| 上海马陆葡萄艺术村 | 168 | 45 | 115 | 52 | ... | 1107 | 1010 |
| 上海环球金融中心观光厅 | 40 | 198 | 82 | 470 | ... | 214 | 67 |
| 上海植物园 | 2425 | 3098 | 4591 | 5753 | ... | 3059 | 4040 |
| 上海南翔景区 | 2050 | 2416 | 2283 | 2659 | ... | 3096 | 3393 |
| 上海徐家汇源景区 | 1529 | 985 | 2589 | 1718 | ... | 1483 | 2233 |
| 上海辰山植物园 | 210 | 165 | 181 | 240 | ... | 1157 | 274 |
| 上海宝山国际民间艺术博览馆景区 | 188 | 166 | 211 | 176 | ... | 4326 | 165 |
| 上海滨海森林公园 | 2 | 12 | 2 | 2 | ... | 2 | 50 |
| 上海田子坊景区 | 900 | 2682 | 1309 | 2335 | ... | 668 | 2238 |
| 上海 M50 创意园 | 909 | 1244 | 574 | 1290 | ... | 798 | 1533 |
| 上海和平公园 | 1242 | 2377 | 1865 | 2841 | ... | 1875 | 2586 |
| 上海醉白池公园 | 109 | 166 | 66 | 156 | ... | 437 | 241 |
| 上海宏泰园 | 45 | 65 | 45 | 68 | ... | 14 | 13 |
| 周浦花海 | 26 | 1 | 27 | 92 | ... | 208 | 1176 |

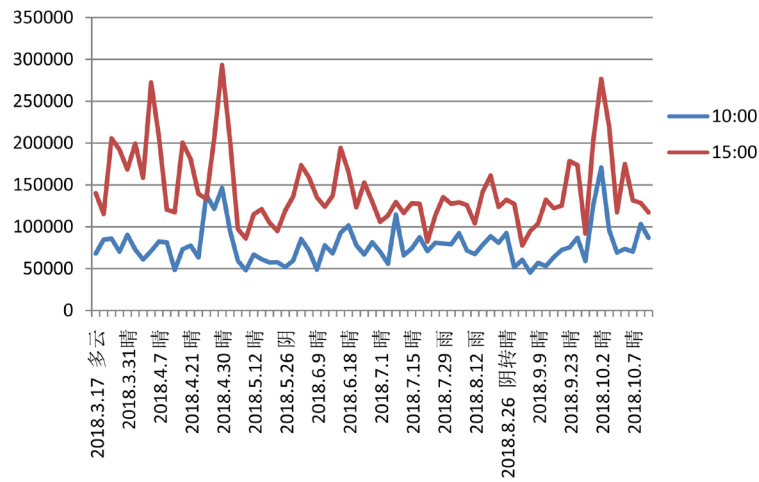


Figure 1. Contrast chart of passenger flow in different periods of holidays
图 1. 节假日不同时间段客流量对比图

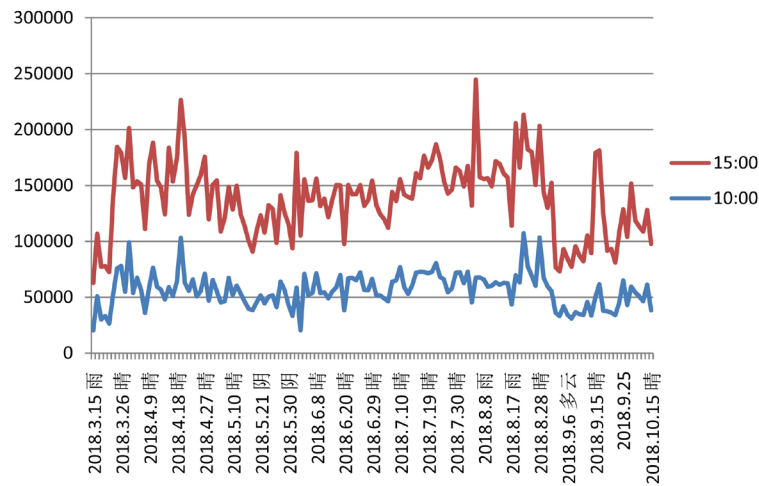


Figure 2. Contrast chart of passenger flow in different periods of non-holidays
图 2. 非节假日不同时间段客流量对比图

| type | 方法 | 均值 | 95%置信均值 | | 标准差 | 95%置信限标准差 | |
|--------|---------------|----------|----------|----------|---------|-----------|---------|
| A | | 77722.6 | 72043.1 | 83402.1 | 23284.3 | 19900.9 | 28064.5 |
| B | | 142476 | 131995 | 152958 | 42971.2 | 36727.2 | 51793.0 |
| 差(1-2) | 汇总 | -64753.4 | -76564.5 | -52942.4 | 34559.2 | 30845.7 | 39297.2 |
| 差(1-2) | Satterthwaite | -64753.4 | -76597.2 | -52909.7 | | | |

| 方法 | 方差 | 自由度 | t值 | Pr > t |
|---------------|-----|--------|--------|---------|
| 汇总 | 等于 | 132 | -10.84 | <.0001 |
| Satterthwaite | 不等于 | 101.68 | -10.84 | <.0001 |

| 方差齐性 | | | | |
|------|-------|-------|------|--------|
| 方法 | 分子自由度 | 分母自由度 | F值 | Pr > F |
| 折叠的F | 66 | 66 | 3.41 | <.0001 |

Figure 3. Testing the difference of passenger flow between 10:00 and 15:00 in holidays
图 3. 节假日 10:00 和 15:00 客流量差异性检验

2.2. 节假日与非节假日景区客流量差异分析

从图 4 所示的 2018 年 3 月 15 日~2018 年 10 月 15 日的景区客流量时序图可以看出, 第一, 在清明节、劳动节、儿童节、端午节、中秋节、国庆节等节假日客流量均出现了不同程度高峰。清明节、端午节、中秋节等节假日有其传统意义, 更多的倾向于在家中度过, 且休息时间都在 1~3 天, 不方便远距离的出行, 所以客流量高峰主要有上海市内游客引起。相比普通节假日而言, 五一和十一期间的客流量高峰明显突出了很多, 这两个假日休息时间长, 除去长期居住在上海的游客外, 客流量高峰主要由外来游客引起。但是其客流量接近平时的两倍, 对于旅游景区来说, 人流过大, 负担太重, 应针对性的进行人流分散和控制。第二, 在 7 月和 8 月份的暑假期间, 客流量整体趋势比非暑假期间高出一截, 而 9 月份后, 客流量出现了较为明显的下降, 说明在上海的游客中, 学生群体也占据了一定比例, 学生的放假和开学都能对客流量带来一定程度的影响。

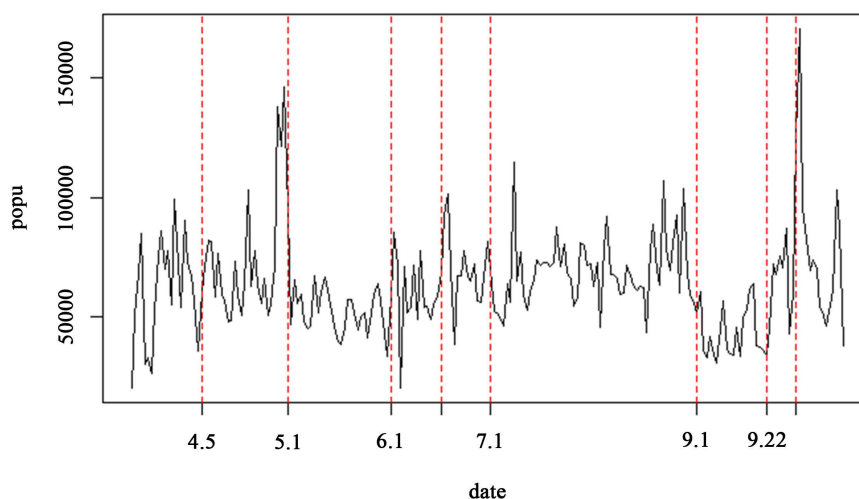


Figure 4. Passenger flow of scenic spots in holiday and non-holiday

图 4. 节假日与非节假日景区客流量

2.3. 基于 K-Means 法对旅游景点聚类分析

K-means 聚类算法是先随机选取 K 个对象作为初始的聚类中心。然后计算每个对象与各个种子聚类中心之间的距离, 把每个对象分配给距离它最近的聚类中心。聚类中心以及分配给它们的对象就代表一个聚类。一旦全部对象都被分配了, 每个聚类的聚类中心会根据聚类中现有的对象被重新计算。这个过程将不断重复直到满足某个终止条件。

本项目对 35 个旅游景点每日 10:00 和 15:00 客流量进行 K-means 聚类分析, 希望找出相似客流量的旅游景点。

因为变量值变化很大, 所以聚类前将数据标准化。如图 5 和图 6 显示从一类到四类变化是, 组内的平方总和有一个明显的下降趋势, 四类之后, 下降的速度减弱, 说明 35 个旅游景点较大概率被聚成四类。

将 35 个旅游聚合成四类, 通过图 7 可以发现上海国际旅游度假区为一类, 上海植物园为一类, 上海野生动物园与东方绿舟为一类, 其余景点为一类。对聚类结果进行猜测, 上海国际旅游度假区和上海植物园每日客流量对比其他景区有显著性差异, 上海野生动物园与东方绿舟每日的客流量也较多, 在第四类的景区中也可以发现上海科技馆与上海动物园分为一类, 猜测这两个景点的主要客流量为学生, 所以客流量的波动趋势较为相似。

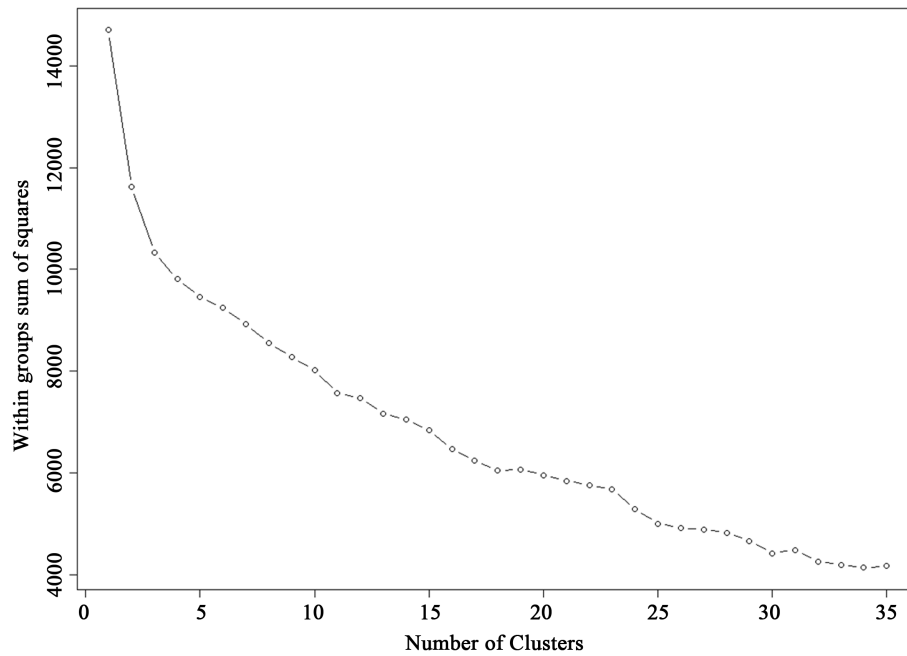


Figure 5. Comparison of the sum of squares in a group and the number of clusters extracted
图 5. 组内的平方和和提取的聚类个数的对比

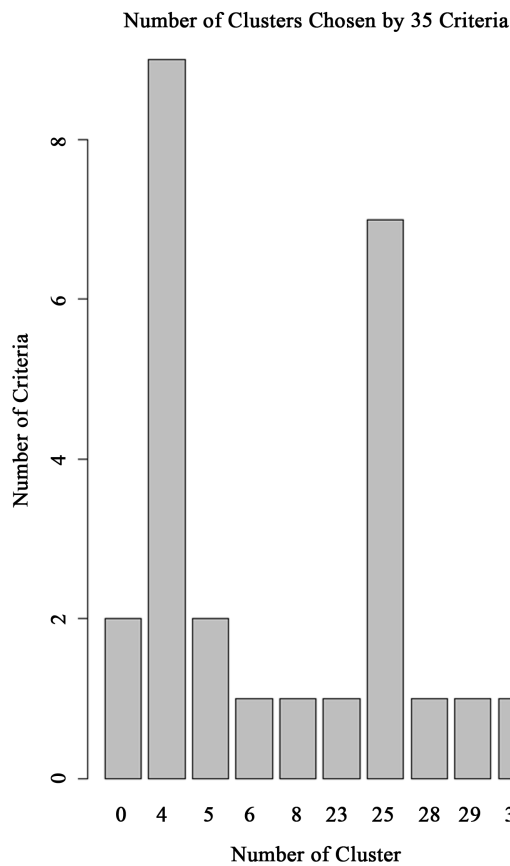


Figure 6. Comparison of the sum of squares in a group and the number of clusters extracted
图 6. 组内的平方和和提取的聚类个数的对比

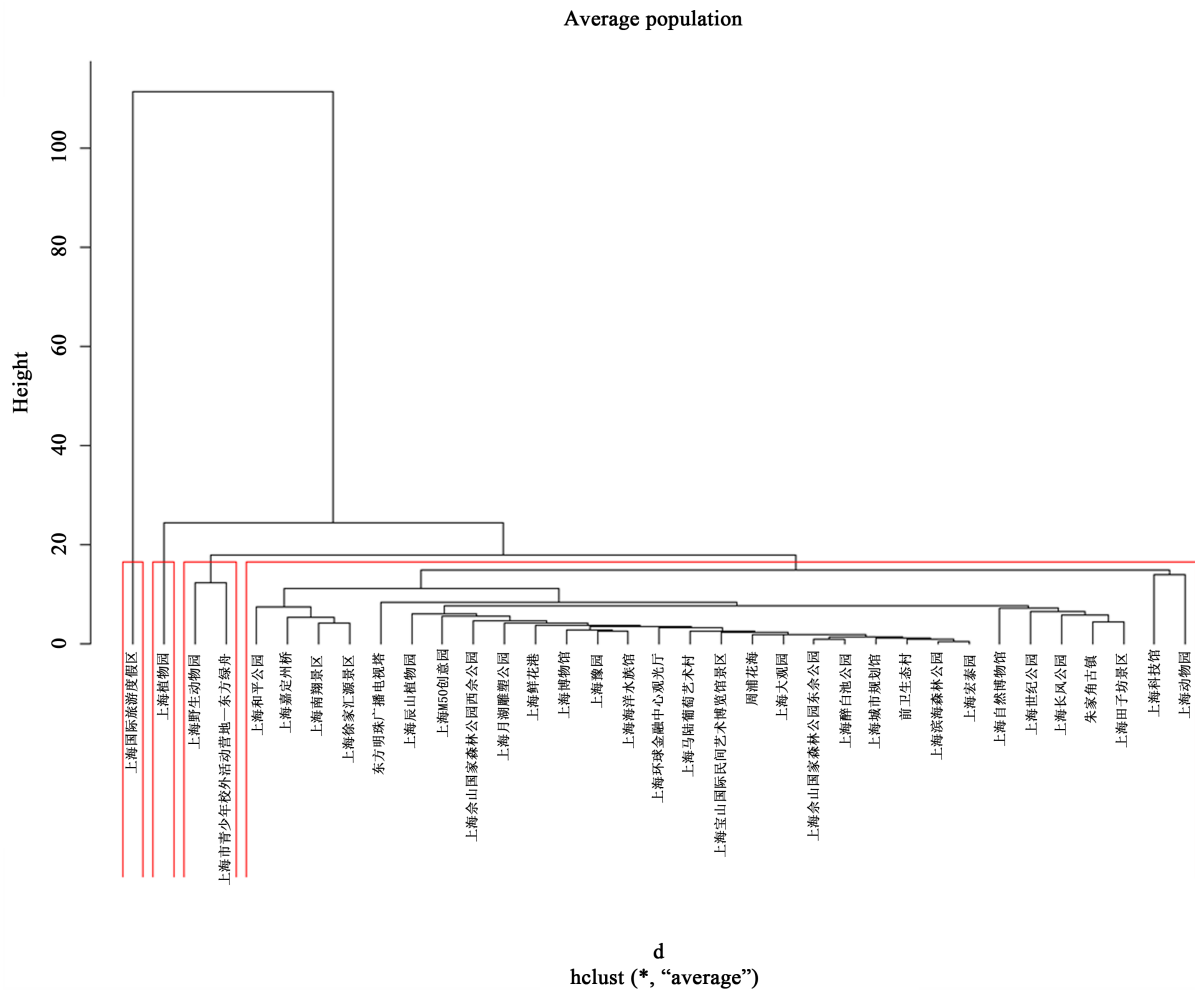


Figure 7. Clustering passenger flow in 35 scenic spots

图 7. 对 35 个旅游景点客流量进行聚类

本文认为此次聚类效果较为不理想的原因在于景点之间的客流量差距过大，几个热门景点客流量一直处于高位而其他景点仍有很大的发展空间。

3. ARIMA 模型

3.1. 关于 ARIMA 模型

ARIMA 模型全称为自回归积分滑动平均模型(Autoregressive Integrated Moving Average Model, 简记 ARIMA), 是由博克斯(Box)和詹金斯(Jenkins)于 70 年代初提出一著名时间序列(Time-series Approach)预测方法, 所以又称为 Box-Jenkins 模型、博克斯 - 詹金斯法。其中 ARIMA (p,d,q)称为差分自回归移动平均模型, AR 是自回归, p 为自回归项; MA 为移动平均, q 为移动平均项数, d 为时间序列成为平稳时所做的差分次数。所谓 ARIMA 模型, 是指将非平稳时间序列转化为平稳时间序列, 然后将因变量仅对它的滞后值以及随机误差项的现值和滞后值进行回归所建立的模型。

ARIMA 模型的基本思想是: 将预测对象随时间推移而形成的数据序列视为一个随机序列, 用一定的数学模型来近似描述这个序列。这个模型一旦被识别后就可以从时间序列的过去值及现在值来预测未来值。

建立 ARIMA 模型的基本程序

1) 根据时间序列的散点图、自相关函数和偏自相关函数图以 ADF 单位根检验其方差、趋势及其季节性变化规律, 对序列的平稳性进行识别。一般来讲, 经济运行的时间序列都不是平稳序列。

2) 对非平稳序列进行平稳化处理。如果数据序列是非平稳的, 并存在一定的增长或下降趋势, 则需要对数据进行差分处理, 如果数据存在异方差, 则需对数据进行技术处理, 直到处理后的数据的自相关函数值和偏相关函数值无显著地异于零。

3) 根据时间序列模型的识别规则, 建立相应的模型。若平稳序列的偏相关函数是截尾的, 而自相关函数是拖尾的, 可断定序列适合 AR 模型; 若平稳序列的偏相关函数是拖尾的, 而自相关函数是截尾的, 则可断定序列适合 MA 模型; 若平稳序列的偏相关函数和自相关函数均是拖尾的, 则序列适合 ARMA 模型。截尾是指时间序列的自相关函数(ACF)或偏自相关函数(PACF)在某阶后均为 0 的性质(比如 AR 的 PACF); 拖尾是 ACF 或 PACF 并不在某阶后均为 0 的性质(比如 AR 的 ACF)。

4) 进行参数估计, 检验是否具有统计意义。

5) 进行假设检验, 诊断残差序列是否为白噪声。

6) 利用模型预测未来值。

3.2. 建立 ARIMA 模型

根据聚类分析的结果, 分别对四类景区建立 ARIMA 时间序列模型进行拟合。

3.2.1. 上海国际旅游度假区

从图 8 中可以看出, 从三月到十月的客流量数据波动较大且不平稳, 其中三月至八月客流量在 40,000 左右波动, 而九月以后, 客流量有明显的下降, 在 30,000 上下波动。对时间序列进行扩展的 Dickey-Fuller 单位根检验, 检验结构结果表明序列不平稳, 存在单位根。由此, 对序列进行差分处理, 差分后的序列明显是平稳的, ADF 检验也拒绝了存在单位根的原假设。

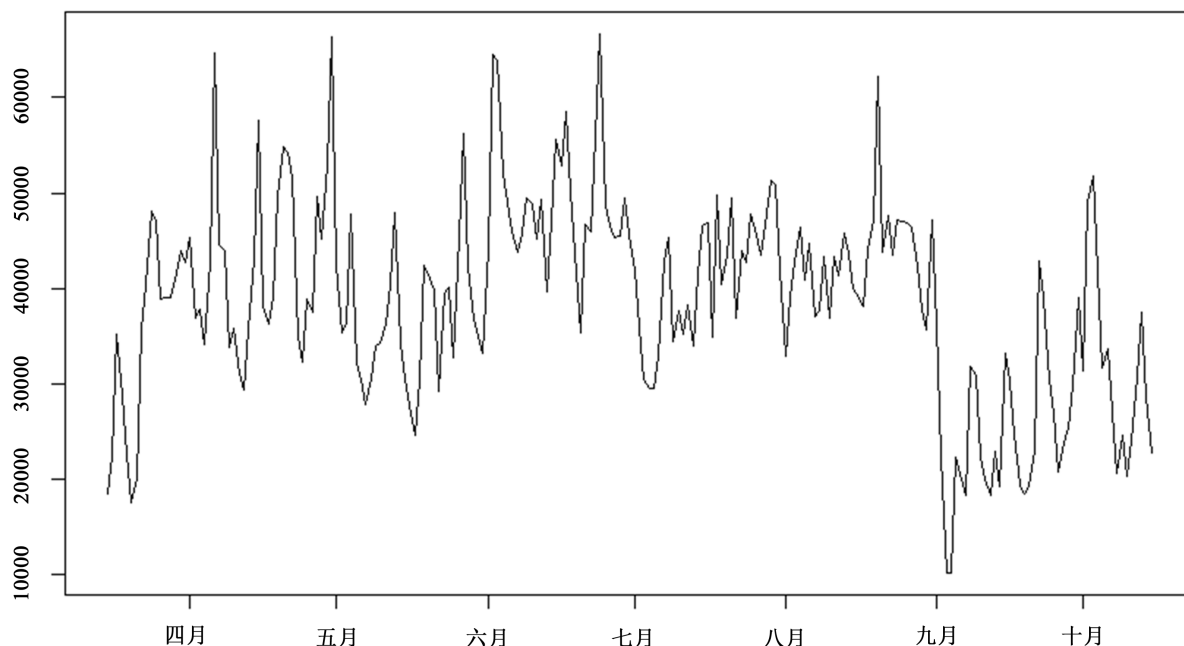


Figure 8. Time series map of passenger flow in Shanghai international tourism resort
图 8. 上海国际旅游度假区客流量时序图

差分后的自相关函数(ACF)图中, 2、3、4 阶存在微弱的显著性, 7、14、18、21、25、28 阶也存在一定的显著性, 序列存在一定的周期。

偏自相关(PACF)图中, 前 2~6 阶是较为显著的, 14 和 18 阶也有一定的显著性。

推广的自相关函数 EACF 识别了一个 ARIMA (3,1,4)模型。

分别用 ARIMA (6,1,4)、ARIMA (3,1,4)以及 ARIMA (1,1,1) × (0,1,1), 周期为 7, 对数据进行拟合, 根据 AIC 准则, ARIMA (1,1,1) × (0,1,1)模型是较为合适的。

拟合的模型为: $(X_t - 0.5572X_{t-1})(X_t - X_{t-1}) = (a_t + 0.8693a_{t-1})(a_t + 0.9014a_{t-7})$

模型检验: 如图 9 所示, 模型的残差检验表明残差是白噪声序列, 建立的模型是合理的。

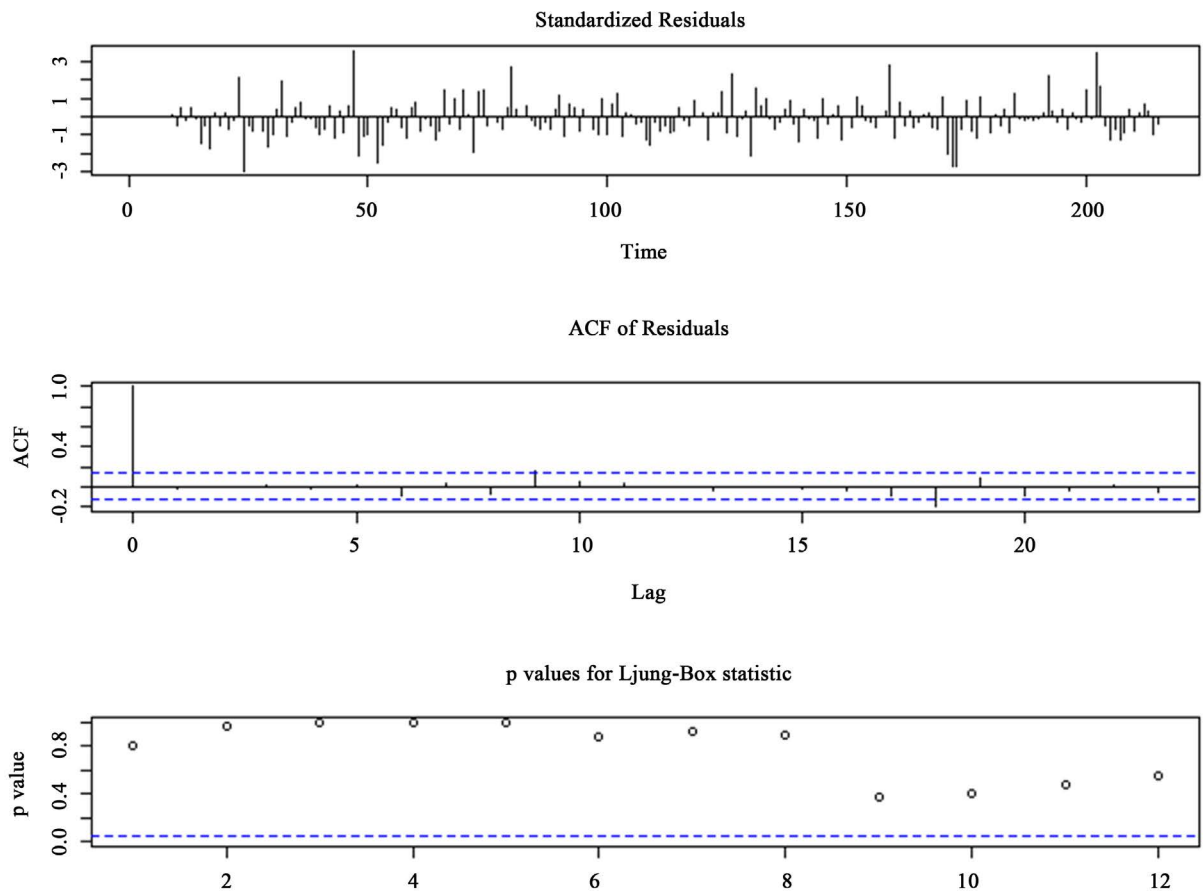


Figure 9. Residual test chart

图 9. 残差检验图

模型预测: 上海国际旅游度假区两周的客流量预测值如表 2 所示。

Table 2. Forecast value of passenger flow in Shanghai international tourism resort

表 2. 上海国际旅游度假区客流量预测值

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|----------|----------|----------|----------|----------|----------|----------|
| 22555.11 | 23037.69 | 23902.67 | 25695.13 | 31606.85 | 30094.33 | 25363.39 |
| 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| 23651.69 | 23311.89 | 23718.63 | 25255.73 | 31025.16 | 29433.36 | 24658.23 |

从建立的模型中可以得出, 上海国际旅游度假区的客流量存在先上升后下降的趋势, 上升的周期大约为一周, 这也是符合现实情况的, 随着周末的到来, 出行的人数相对平时会有所上升。另外, 客流量在时间上存在自相关性, 表现为前一天的客流量对今天的客流量存在影响。

3.2.2. 上海植物园

如图 10 所示, 上海植物园的客流量在三四五月份较高, 六月至八月份客流量下降, 且比较平稳, 波动不大, 九月以后又出现了上升趋势。整体上来看, 序列是不平稳的, ADF 单位根检验的结果也表明序列不平稳, 不能直接建立模型, 所以对数据进行差分处理。

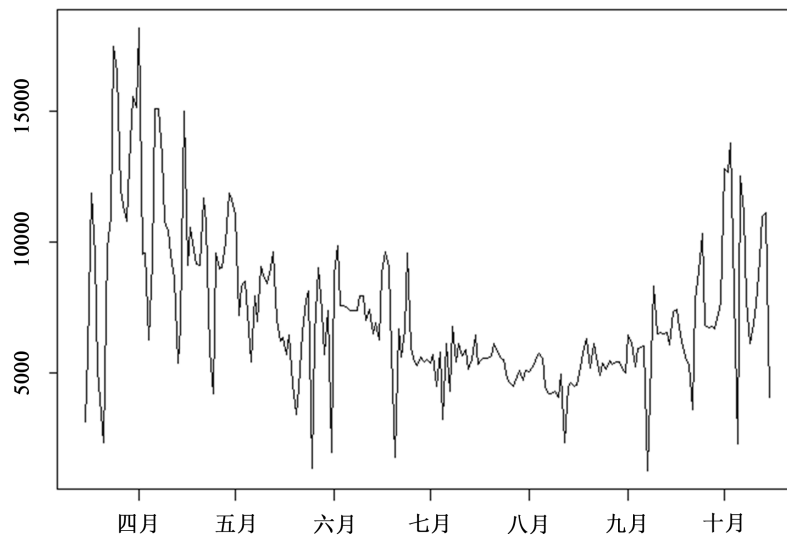


Figure 10. Time series chart of passenger flow in Shanghai botanical garden
图 10. 上海植物园客流量时序图

如图 11 所示, 差分后的数据是平稳的。且可以从差分的图中看出, 上海植物园的客流量在 3 月~6 月以及 10 月的波动性较大, 7~9 月波动性小, 较为稳定。

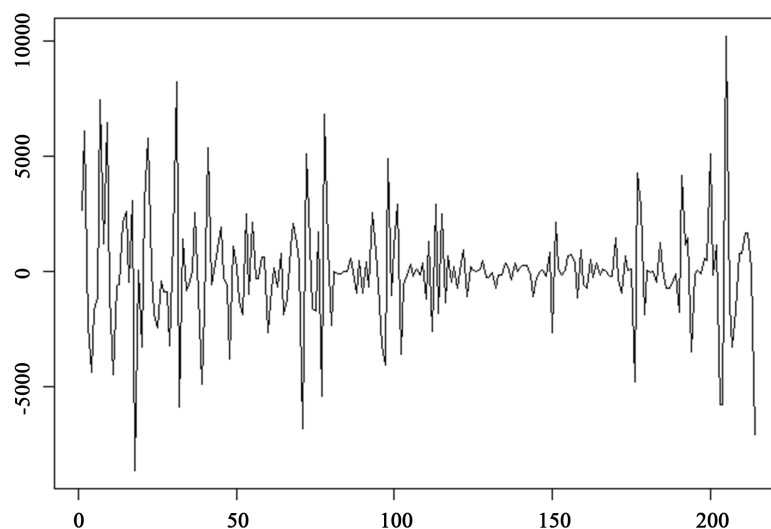


Figure 11. Differenced sequence diagram
图 11. 差分后的时序图

从自相关函数(ACF)中可以看出, 前三阶存在较为明显的显著性, 14 阶也有一些显著性, 但之后没有明显的周期性。

从偏自相关(PACF)图中可以看出, 前五阶存在较为明显的显著性。

扩展的自相关函数 EACF 识别了一个 ARIMA (2,1,2)模型。

根据 AIC 准则, ARIMA (2,1,2)模型较为合适。

拟合的模型: $X_t - 1.0628X_{t-1} + 0.5266X_{t-2} = a_t + 1.4771a_{t-1} - 0.6242a_{t-2}$

模型检验: 如图 12 所示, 模型的残差检验表明残差是白噪声序列, 建立的模型是合理的。

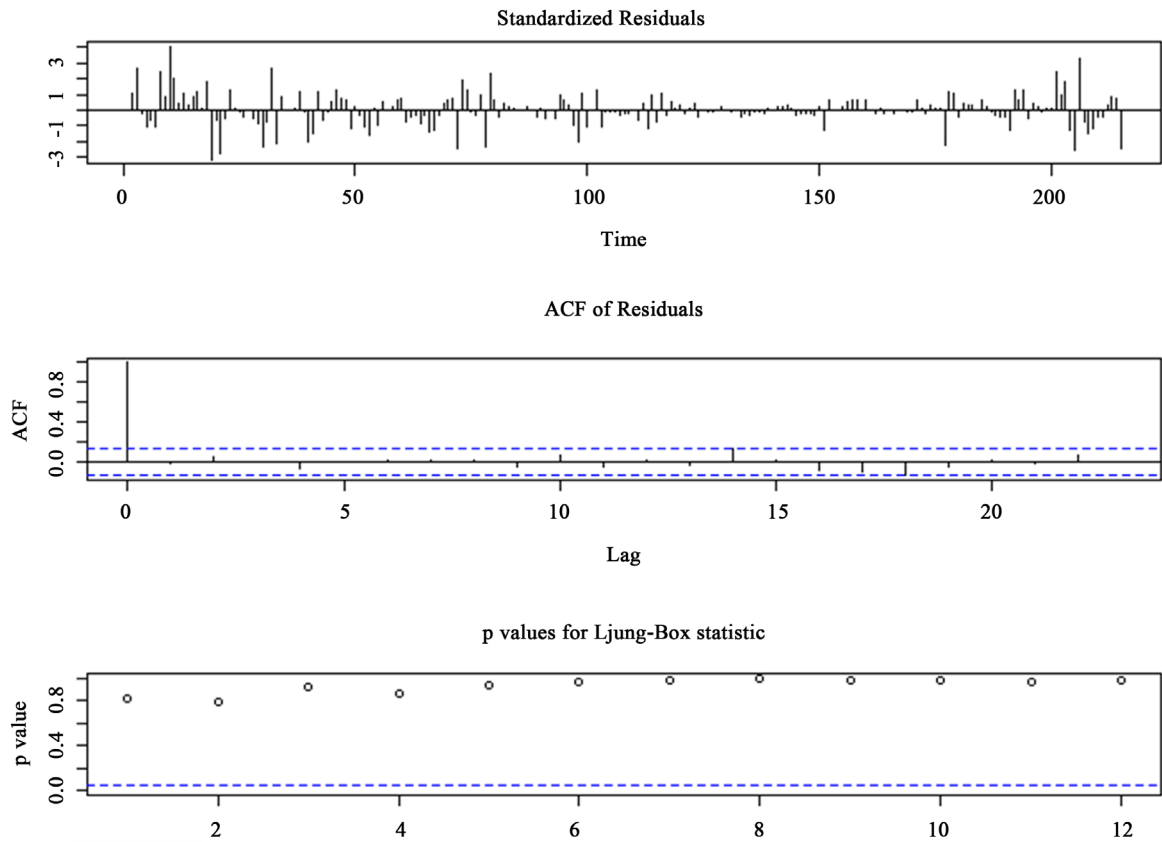


Figure 12. Model check chart

图 12. 模型检验图

模型预测: 上海植物园两周的客流量预测值如表 3 所示。

Table 3. Forecast value of passenger flow in Shanghai botanical garden

表 3. 上海植物园客流量预测值

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|----------|----------|----------|----------|----------|----------|----------|
| 5161.777 | 6788.321 | 7926.305 | 8279.202 | 8054.976 | 7630.812 | 7298.082 |
| 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| 7167.823 | 7204.605 | 7312.297 | 7407.386 | 7451.735 | 7448.794 | 7422.312 |

根据建立的模型可以得出结论, 上海植物园的客流量在时间上的自相关性表现为前两天的客流量数据对今天的客流量存在影响, 且前两天的客流量数据的波动也对今天的客流量存在影响。

3.2.3. 上海野生动物园和东方绿洲

如图 13 所示, 上海野生动物园和东方绿舟两个景区的客流量变化确实较为相似, 其中, 东方绿舟的客流量稍微高一些, 选择东方绿舟的客流量建立模型。

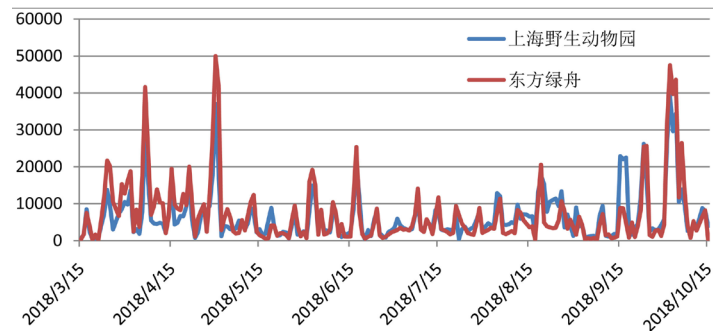


Figure 13. Time series diagram of passenger flow in Shanghai wildlife park and oriental green boat
图 13. 上海野生动物园和东方绿舟客流量时序图

东方绿舟的客流量除了节假日出现较高的峰值外, 处于较为平稳的变动之中, 对序列的平稳性进行检验, ADF 检验也拒绝了存在单位根的原假设, 序列是平稳的, 可以直接建立 ARIMA 模型。

自相关函数(ACF)图中, 前两阶是显著的。

偏偏相关函数(PACF)图中第一阶是非常显著的, 第二阶也有微弱的显著性。

拟合 ARIMA (1,1,2)和 ARIMA (2,1,2)模型, 对比两个模型的 AIC, ARIMA (1,1,2)模型更为显著。

拟合的模型: $X_t - 0.4312X_{t-1} = a_t + 0.6676a_{t-1} + 0.3004a_{t-2}$

模型检验: 如图 14 所示, 模型的残差检验表明残差是白噪声序列, 建立的模型是合理的。

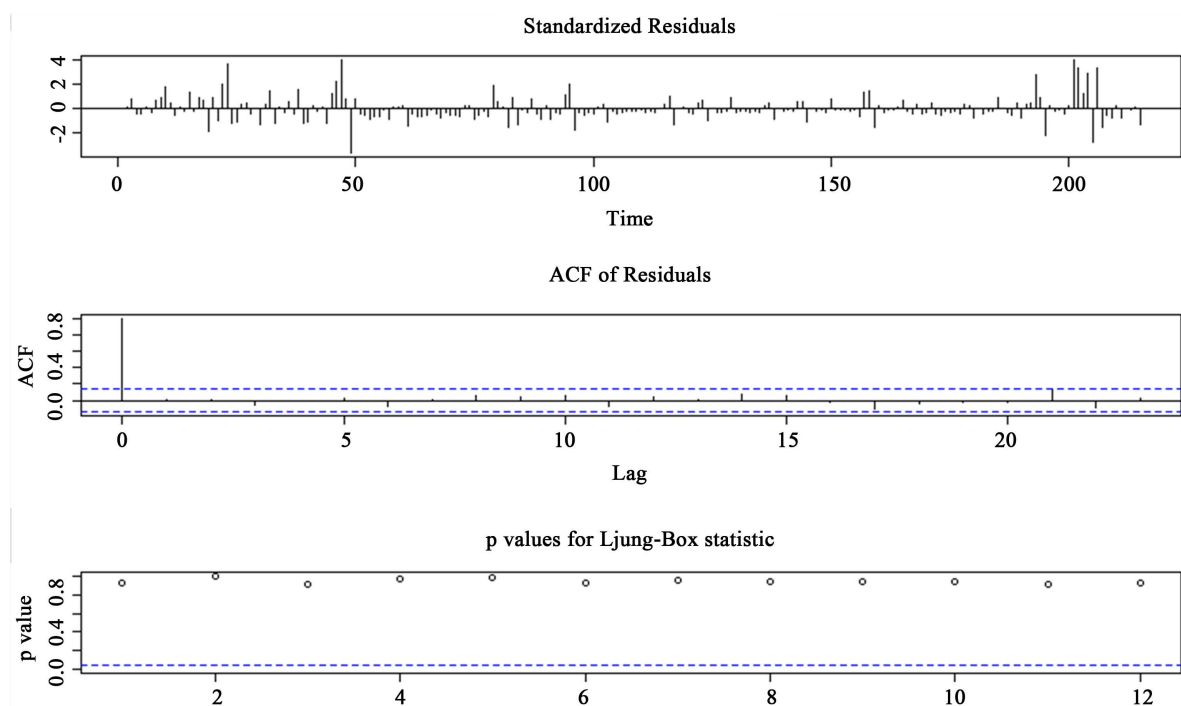


Figure 14. Model check chart
图 14. 模型检验图

模型预测：上海野生动物园和东方绿舟两周的客流量预测值如表 4 所示。

Table 4. Forecast value of passenger flow of Shanghai wildlife park and oriental green boat
表 4. 上海野生动物园和东方绿舟客流量预测值

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|----------|----------|----------|----------|----------|----------|----------|
| 2011.587 | 5348.393 | 6787.235 | 7407.668 | 7675.201 | 7790.562 | 7840.306 |
| 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| 7861.756 | 7871.005 | 7874.994 | 7876.713 | 7877.455 | 7877.775 | 7877.913 |

根据模型可以得出结论，上海野生动物园和东方绿舟这一类的景区的客流量波动较为平稳，整体的客流量没有明显的上升趋势，一直在稳定的水平上波动。时间上的自相关性表现为前一天的客流量对今天的客流量存在影响，前两天的客流量的扰动对今天的客流量的扰动存在影响。

3.2.4. 其他类景区

观察其他景区的数据及趋势后，选择趋势较为明显且客流量相对较高的徐家汇源景区建立模型。如图 15 所示，徐家汇源景区的客流量出现了较多的峰值，但整体上在一个较为平稳的趋势上波动，ADF 单位根检验的结果也证明序列是平稳的，可以直接建立 ARIMA 模型。

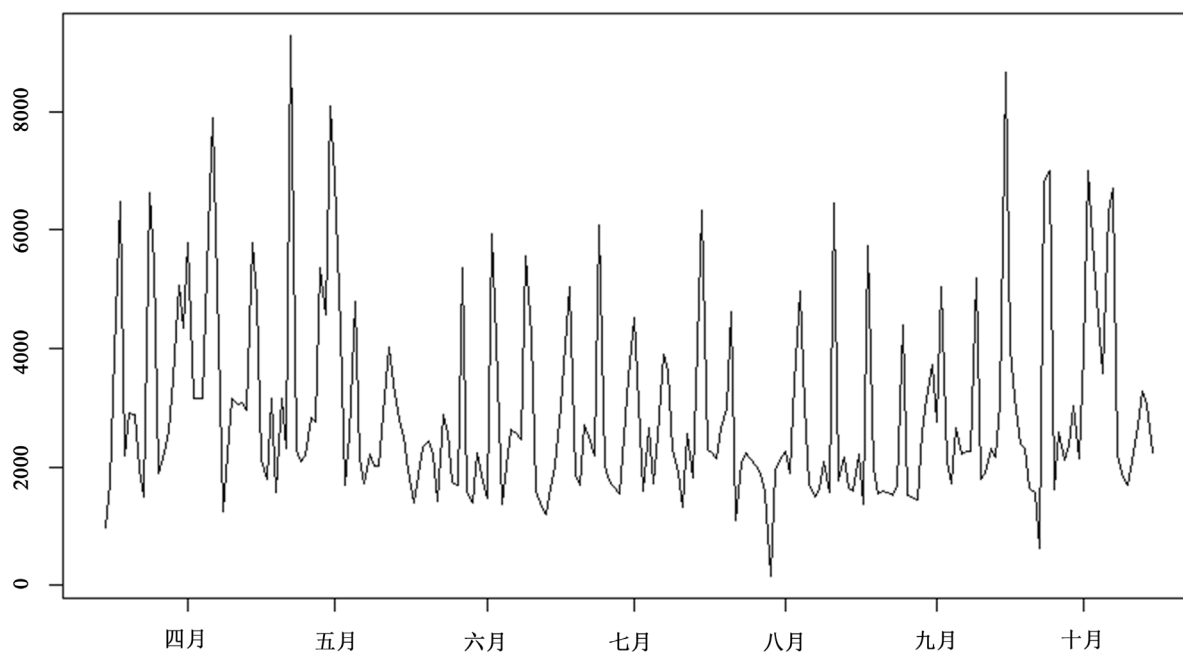


Figure 15. Xujiahuiyuan scenic area
图 15. 徐家汇源景区

从自相关(ACF)图中，序列存在明显的周期性。

偏自相关(PACF)图中，第一阶较为显著。

建立 ARIMA (1,0,0) × (0,1,1)模型，周期为 7。

拟合的模型： $X_t - 0.2528X_{t-1} = X_t + X_{t-7}$

模型检验：如图 16 所示，模型的残差检验表明残差是白噪声序列，建立的模型是合理的。

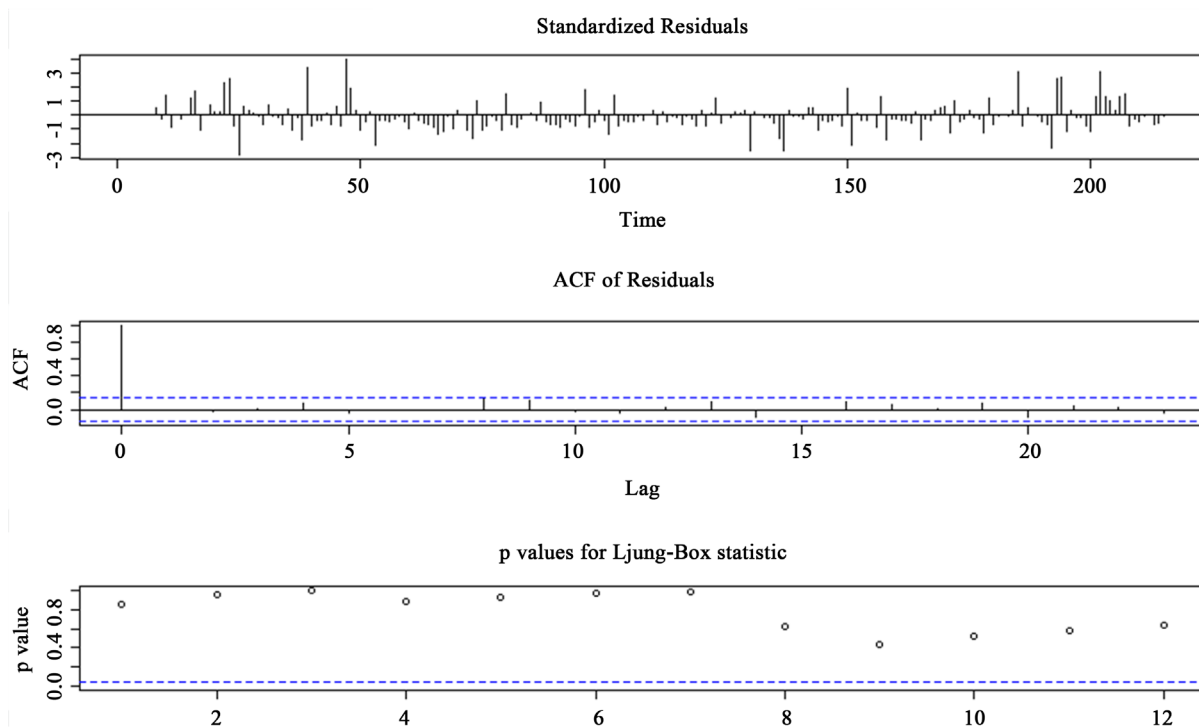


Figure 16. Model check chart

图 16. 模型检验图

模型预测：其他类景区两周的客流量预测值如表 5 所示。

Table 5. Forecast value of passenger flow in other scenic spots

表 5. 其他类景区客流量预测值

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|----------|----------|----------|----------|----------|----------|----------|
| 2219.120 | 2385.845 | 2335.806 | 2640.914 | 4167.536 | 4081.973 | 2625.427 |
| 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| 2318.328 | 2410.925 | 2342.147 | 2642.517 | 4167.941 | 4082.076 | 2625.453 |

根据模型的拟合结果可以得出结论，徐家汇源等景区的客流量数据波动是较为平稳的，整体上的在一个稳定的水平上，节假日会出现一定的波动，但几个月内没有明显的客流量的上升趋势。在时间上的自相关性表现为前一天的客流量对今天的客流量存在影响，且存在周期为 7 天的影响，即一周以前的客流量数据对今天的客流量数据有一定程度上的影响。

4. 结论与建议

通过上述分析可以得出上海旅游景区每日 10:00 和 15:00 客流量存在显著性差异且 15:00 客流量普遍高于 10:00 客流量。节假日中五一和十一黄金周景区客流量较平日出现激增，双休日相较于工作日景区客流量有小幅上升，在 9 月份景区客流量也呈现下降趋势。通过时间序列分析，上海市 3A 及以上景区客流量在时间上存在相互影响，滞后周期为 1。

可以得出，相对于早上，游客更多的会选择下午的时间段到景区游玩，另外，导致下午景区客流量高于上午的原因可能是游客在景区停留的时间较长，所以到下午，景区的客流量会越来越多。同时，景

区的客流量并不会因为季节的变化而带来巨大的波动,整体上一年中3月到10月的客流量没有太大变化,景区客流量会出现爆炸的时间主要为五一小长假和十一黄金周期间。

针对以上问题,首先,各景区可以为游客规划合理的游玩路线,避免出现游客聚集现象,同时合理的路线也可以减少游玩时间,可以减轻景区下午时间段客流量较大的问题。其次,各景区可以利用 ARIMA 模型预测未来的客流量趋势,提前制定出合理的控制方案。另外,针对“五一”小长假和“十一”黄金周期间加强对景区人流的控制,避免出现旅游容量过饱和,降低游客的游玩体验。

同时,通过 K-means 聚类算法可以得出上海少数几个热门景点如上海国际旅游度假区、上海植物园、上海野生动物园和东方绿舟一直占据了绝大部分客流量,而其他景点客流量依旧处于较低水平,如果上海想要进一步可持续发展旅游业,可以针对其他旅游景点进行特色宣传,促进其他景点的旅游业发展。特别在五一小长假和十一黄金周期间,可以提前加强其他客流量较少景点的宣传,促使客流量分散到不同景区以防止几个热门景点出现客流量饱和的情况。

基于 ARIMA 时间序列模型可以发现四类景区的客流量存在不同的波动性和自相关性。上海国际旅游度假区的客流量数据在几个月内都存在较大的波动,且波动存在周期性。上海植物园的客流量在3~6月和10月以后较高,波动性也较大,而7~9月的客流量相对较低,波动也较为平稳,时间序列的影响上也不存在明显的周期性。上海野生动物园和东方绿舟的客流量除清明、五一小长假、十一黄金周等会出现明显的高峰,其余时间的波动都较为平稳,没有明显的上升趋势,时间序列上的影响表现为前两天的客流量对当下的客流量存在影响。其他景区在3~5月和10月份的客流量波动较大,其余几个月也有较弱的波动,波动表现出周期性。模型检验的结果表明时间序列模型能够较好的拟合客流量数据,建立的 ARIMA 模型能够针对不同的景区进行客流量预测分析,在达到饱和点之前及时进行控制。

本文通过描述性统计分析及时间序列建模的方法,图形与模型并存,直观科学地分析了客流量的分布情况并预测了景区客流量的未来趋势。但由于项目时间原因,收集的数据时间线上不够长,且部分景区的客流量数据更新不及时,导致收集的景区数据覆盖不全面。

参考文献

- [1] 汪宇明, 赵中华. 基于上海案例的大都市旅游容量及承载力研究[J]. 中国人口·资源与环境, 2007, 17(5): 118-122.
- [2] 舒锡慧, 罗茹. 上海旅游容量调查实证分析[J]. 大众文艺, 2011(24): 354-356.
- [3] 刘春济, 高静. 上海市旅游景区(点)客流时空分布特征及成因分析[J]. 上海师范大学学报(自然科学版), 2006(5): 93-100.
- [4] 黄筱焯. 上海都市旅游景点可容纳量研究[D]: [硕士学位论文]. 上海: 上海师范大学, 2005.
- [5] 陈珊, 申世昌. ARIMA 模型对我国旅游业可持续发展的实证研究[J]. 齐齐哈尔大学学报(自然科学版), 2017, 33(2): 84-88.
- [6] 刘燕玉, 龙金萍. 基于 ARIMA 模型的江口县旅游人数实证分析[J]. 统计与管理, 2016(6): 36-37.
- [7] 赵玥. 时间序列分析方法在景点人数预测问题中的应用[D]: [硕士学位论文]. 哈尔滨: 黑龙江大学, 2018.

知网检索的两种方式：

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择：[ISSN]，输入期刊 ISSN：2325-2251，即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入，输入文章标题，即可查询

投稿请点击：<http://www.hanspub.org/Submission.aspx>

期刊邮箱：sa@hanspub.org