

基于灰色GM(1, 1)模型的新冠肺炎感染人数预测研究

魏 齐¹, 王 丽^{2*}

¹内江职业技术学院通识教育与公共服务学院, 四川 内江

²内江职业技术学院智能建造学院, 四川 内江

收稿日期: 2024年7月5日; 录用日期: 2024年7月26日; 发布日期: 2024年8月8日

摘 要

本文以2020年2月6日至2020年2月19日日本“钻石公主号”邮轮上新冠肺炎感染人数为基础, 首先建立灰色预测模型研究感染人数的趋势情况。接着, 运用时间序列分析理论讨论灰色预测模型的残差是否为白噪声序列, 以进一步建立相应的时间序列模型, 从而提高灰色预测模型的精度。计算结果表明, 该方法能有效地研究此封闭空间中的新冠肺炎确诊人数。

关键词

新冠肺炎, 灰色预测模型, 白噪声检验, 残差

Study on the Prediction of the Number of Confirmed Cases of COVID-19 Based on the Grey GM(1, 1) Model

Qi Wei¹, Li Wang^{2*}

¹College of General Education and Public Service, Neijiang Vocational & Technical College, Neijiang Sichuan

²College of Intelligent Construction, Neijiang Vocational & Technical College, Neijiang Sichuan

Received: Jul. 5th, 2024; accepted: Jul. 26th, 2024; published: Aug. 8th, 2024

Abstract

Based on the number of confirmed cases of COVID-19 in Japan “Diamond Princess” cruises from

*通讯作者。

February 6, 2020 to February 19, 2020, we first establish the grey prediction model to study the trend of the number of confirmed cases. Then, the time series analysis theory is applied to discuss whether residuals of the grey prediction model are white noise sequence, so as to further build the corresponding time series models to improve the accuracy of the grey prediction model. The results show that the number of confirmed cases of COVID-19 in such a closed space can be effectively diagnosed by this method.

Keywords

Novel Coronavirus Pneumonia, Grey Prediction Model, White Noise Test, Residual Error

Copyright © 2024 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

灰色预测模型是我国学者邓聚龙教授提出的, 其中 GM(1, 1) [1]模型是灰色预测模型的基础, 其主要内容有白化方程和对应的差分方程。通过白化方程可以推导出模型的时间函数和还原值的表达式, 通过差分方程可以给出模型参数的表达式, 并在此基础上对所研究的问题进行拟合预测。经过几十年的发展, 许多作者围绕 GM(1, 1)模型开展了一系列的研究工作, 既有理论上的研究, 也有应用上的研究, 相关研究工作具体见文献[2]-[7]。另一方面, 从统计学教材[8] [9]可知, 当数据的变化表现出稳定的下降或者增长时, 可用线性模型来对数据进行建模。当数据的变化表现出非线性时, 可建立合适的趋势模型。在本文中, 我们将对灰色预测模型的残差进一步分析, 以确定是否有进一步对残差建模的必要, 以提高灰色预测模型的精度和准度。

“钻石公主号”(Diamond Princess), 位列全球十五大最豪华邮轮之一。该轮是驰名世界的邮轮品牌“公主号”系列船队中体积最庞大、设施最完善的世界顶级豪华邮轮之一, 犹如一座海上的五星级酒店。在此次全球新型冠状病毒疫情下, “钻石公主号”邮轮同样备受关注。此邮轮搭载了 2666 名乘客和 1045 名船员, 在 2020 年 1 月 25 号时, 该邮轮上一名 80 岁的乘客在香港下船后, 于 2020 年 2 月 1 日确诊为新型冠状病毒肺炎。受此消息影响, “钻石公主号”提前返回横滨, 并于 2 月 3 日晚间开始接受日本厚生劳动省的检疫。2020 年 2 月 6 日, 新型冠状病毒确诊人数达到了 20 人, 2020 年 2 月 19 日, “钻石公主号”邮轮滞留乘客正式开始下船, 当日就约有 500 名乘客下船。在本文中, 我们将利用灰色预测模型和时间序列分析方法讨论感染人数的变化趋势, 从而为病毒传播提供一些可供参考的依据。

本文的结构具体安排如下: 第 2 节给出单变量灰色 GM(1, 1)模型的具体建模步骤, 以及判断白噪声的相关准则。第 3 节以 2020 年 2 月 6 日至 2020 年 2 月 19 日日本“钻石公主号”邮轮上新型冠状病毒感染人数为基础, 建立相应的灰色预测模型, 并对残差进行白噪声检验。第 4 节给出了本文的结论。

2. 灰色 GM(1, 1)模型和白噪声检验

本小节首先将给出灰色预测模型的建模步骤、模型的响应函数和还原值、参数的最小二乘估计等。接着, 给出判断残差序列是否为白噪声的检验准则。

2.1. 灰色 GM(1, 1)模型

设原始数据系列为 $X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$, 其中数据 $x^{(0)}(k) \geq 0, k = 1, 2, \dots, n$, 则称

$X^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n))$ 为序列 $X^{(0)}$ 的一次累加生成序列, 其中 $x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i), k = 1, 2, \dots, n$, 则称

$$\frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = b, \quad (1)$$

为灰色 GM(1, 1) 模型的白化方程。进一步, 称

$$x^{(0)}(k) + az^{(1)}(k) = b, \quad (2)$$

为灰色 GM(1, 1) 模型的差分形式, 其中背景值 $z^{(1)}(k) = 0.5x^{(1)}(k-1) + 0.5x^{(1)}(k)$ 。

由灰色 GM(1, 1) 模型的白化方程, 运用一阶非齐次常微分方程理论, 可以得到模型的响应函数、还原值的表达式分别为:

$$\hat{x}^{(1)}(k) = \left[x^{(0)}(1) - \frac{b}{a} \right] e^{-a(k-1)} + \frac{b}{a}, k = 1, 2, \dots, \quad (3)$$

$$\hat{x}^{(0)}(k) = \frac{e^a - 1}{a} \left(b - x^{(0)}(1) \right) e^{-a(k-1)}, k = 1, 2, \dots. \quad (4)$$

由灰色 GM(1, 1) 模型的差分形式, 结合最小二乘方法可以推导模型参数的表达式为下列形式:

$$\begin{pmatrix} a \\ b \end{pmatrix} = (B^T B)^{-1} B^T Y, \quad (5)$$

$$\text{其中, } B = \begin{pmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(m) & 1 \end{pmatrix}, Y = \begin{pmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(m) \end{pmatrix}.$$

一旦给定原始数据序列 $X^{(0)}$ 和建模个数 m 的值, 就可以通过上述的表达式(3)~(5)完成建模和预测。

2.2. 评价模型精度的准则

在灰色预测模型研究中, 有许多评价模型精度的准则, 其中百分比误差、均方根误差是常用的评价准则, 其表达式如下:

$$\text{百分比误差: } APE(k) = \left| 1 - \frac{\hat{x}(k)}{x(k)} \right| \times 100\%, k = 1, 2, \dots, n. \quad (6)$$

$$\text{均方根误差: } RMSPE = \sqrt{\frac{1}{r-l+1} \sum_{k=l}^r \left(1 - \frac{\hat{x}(k)}{x(k)} \right)^2} \times 100\%. \quad (7)$$

其中, $l=1, r=m$ 表示建模误差, $l=m+1, r=n$ 表示拟合误差, $l=1, r=n$ 表示总的误差。

2.3. 白噪声检验

由文献[9]知道, 白噪声序列是原始序列信息充分提取后剩下的残差序列, 表明所建立的模型能够很好地提取序列信息。本小节就残差序列是否为白噪声给出三种常见的判断准则。

自相关系数判断法: 如果灰色预测模型的残差序列的自相关系数随延迟期数的增加迅速或者突然衰减到零的附近小幅波动, 则可推断出残差序列是平稳白噪声序列。

时序图判断法: 对灰色预测模型的残差序列 $APE(k), k = 1, 2, \dots, n$ 利用统计软件作出时序图, 若序列

值围绕一个常数随机小幅波动, 且没有明显的周期性和趋势性, 则可推断出残差序列为平稳白噪声序列。

P 值法: 对于残差序列 APE, 记样本的自相关系数为 $\hat{\rho}_l, l=1, 2, \dots, \nu$, ν 为指定的延迟期数。统计量 LB 定义为:

$$LB(h, \nu) = h(h+2) \sum_{k=1}^{\nu} \frac{\hat{\rho}_k^2}{h-k} \sim \chi^2(\nu). \quad (8)$$

原假设 H_0 、备择假设 H_1 分别为:

$$H_0: \rho_1 = \rho_2 = \dots = \rho_\nu = 0, \forall \nu \geq 1; \quad H_1: \exists 1 \leq k \leq \nu, \rho_k \neq 0, \forall \nu \geq 1. \quad (9)$$

给定检验水平 α (本文取 $\alpha = 0.05$), 在原假设 H_0 下, 计算 $P\{LB(h, \nu) > \chi_\alpha^2(\nu)\}$ 的值, 当 $P\{LB(h, \nu) > \chi_\alpha^2(\nu)\} > \alpha$ 接受 H_0 , 序列 APE 为白噪声序列; 当 $P\{LB(h, \nu) > \chi_\alpha^2(\nu)\} \leq \alpha$ 拒绝 H_1 , 序列 APE 为非白噪声序列, 需进一步为残差序列建模分析, 且将计算的结果融合到灰色预测模型的计算值上, 进而得到最终的结果。

3. 新冠肺炎感染人数趋势分析

在本节中, 我们将以 2020 年 2 月 6 日至 2020 年 2 月 19 日日本“钻石公主号”邮轮上新冠肺炎感染人数为基础, 来建立灰色 GM(1, 1) 预测模型, 原始数据来源于网站公布的数据

(https://m.sohu.com/a/374965922_120359141), 具体见表 1。

Table 1. Number of confirmed cases of COVID-19 in Japan “Diamond Princess” cruises

表 1. 日本“钻石公主号”邮轮上新冠肺炎感染人数

日期	数据	日期	数据	日期	数据
2020 年 2 月 6 日	20	2020 年 2 月 11 日	135	2020 年 2 月 16 日	355
2020 年 2 月 7 日	61	2020 年 2 月 12 日	174	2020 年 2 月 17 日	454
2020 年 2 月 8 日	64	2020 年 2 月 13 日	218	2020 年 2 月 18 日	542
2020 年 2 月 9 日	70	2020 年 2 月 14 日	218	2020 年 2 月 19 日	621
2020 年 2 月 10 日	135	2020 年 2 月 15 日	285		

对表 1 的数据分成两部分, 其中前 11 个数据(2020 年 2 月 6 日~2020 年 2 月 16 日)用来建立灰色预测模型, 后 3 个数据(2020 年 2 月 17 日~2020 年 2 月 19 日)用来检验模型的精度。编写数值计算程序, 并利用 2.1 节公式(3)~(5)给出模型相关参数的值和表达式。经计算得到 $a = -0.190662, b = 53.966605$, 于是模型的响应函数为:

$$\hat{x}^{(1)}(k) = -263.049033e^{0.190662(k-1)} - 283.049033, k = 1, 2, \dots, \quad (10)$$

$$\hat{x}^{(0)}(k) = \frac{1 - e^{-0.190662}}{0.190662} \times 33.966605e^{0.190662(k-1)}, k = 1, 2, \dots. \quad (11)$$

通过计算得到相应的数值结果见表 2 和图 1。

从图 1 看出, 灰色 GM(1, 1) 的计算结果与原始数据拟合的还是比较好的。从表 2 看出, 建模的百分比误差最大值为 33.1509%, 预测的百分比误差最大值为 5.6346%, 其他的还是比较小的。另外, 均方根的建模、预测和总的误差分别为 14.2338%、4.1522% 和 12.6422%。总的来说, 还是可以接受的结果。但是, 灰色 GM(1, 1) 模型是否将感染病人的数据信息提取出来了呢? 针对这一问题, 我们将继续对其残差序列进行分析, 基于 Matlab 和 EViews 软件计算得到想要的结果, 具体见图 2 和表 3。

Table 2. Computational results of the confirmed cases of COVID-19 in the Japan “Diamond Princess” cruises

表 2. 日本 “钻石公主号” 邮轮上新冠肺炎感染人数的计算结果

日期	原始数据	灰色 GM(1, 1)模型	APE (%)
2020 年 2 月 6 日	20	20.0000	0.0000
2020 年 2 月 7 日	61	63.6555	4.3532
2020 年 2 月 8 日	64	77.0263	20.3536
2020 年 2 月 9 日	70	93.2056	33.1509
2020 年 2 月 10 日	135	112.7835	16.4567
2020 年 2 月 11 日	135	136.4737	1.0916
2020 年 2 月 12 日	174	165.1400	5.0920
2020 年 2 月 13 日	218	199.8276	8.3360
2020 年 2 月 14 日	218	241.8014	10.9181
2020 年 2 月 15 日	285	292.5918	2.6638
2020 年 2 月 16 日	355	354.0507	0.2674
2020 年 2 月 17 日	454	428.4190	5.6346
2020 年 2 月 18 日	542	518.4085	4.3527
2020 年 2 月 19 日	621	627.3002	1.0145
RMSPE (%)			14.2338
			4.1522
			12.6422

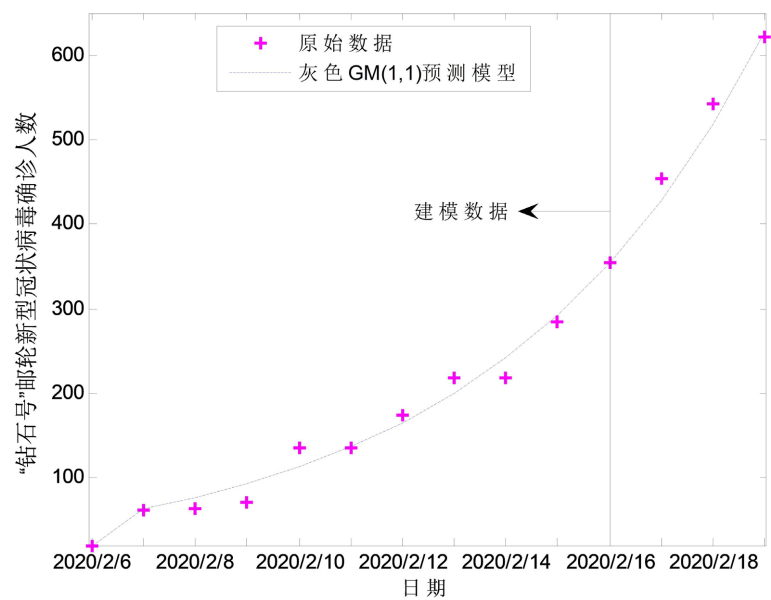


Figure 1. Diagram of computational results of the confirmed cases of COVID-19 in Japan “Diamond Princess” cruises
图 1. 日本 “钻石公主号” 邮轮上新冠肺炎感染人数的计算结果图

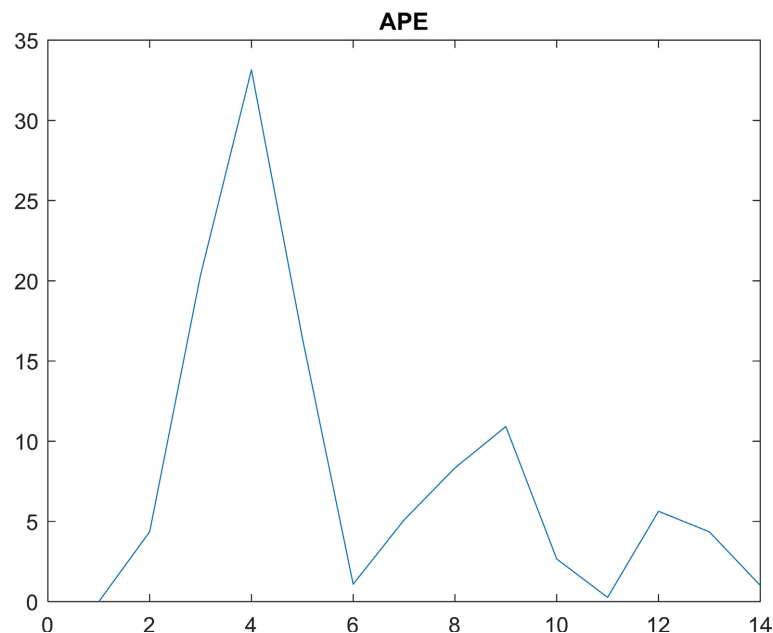


Figure 2. Time series diagram of the residual APE
图 2. 残差 APE 的时序图

从图 2 可以看出，除去第 3、4、5 个值变化较大，其余的残差值波动幅度很小。进一步，残差序列也没有表现出周期性、趋势性。所以，从时序图可以判断灰色预测模型的残差序列为平稳白噪声序列。

Table 3. Correlation coefficient table of APE
表 3. APE 的相关系数表

自相关性	偏相关性		AC	PAC	Q-Stat	Prob
. *** .	. *** .	1	0.464	0.464	3.4936	0.062
. * .	**** .	2	-0.16	-0.477	3.9464	0.139
. * .	. *** .	3	-0.111	0.362	4.1875	0.242
. * .	. * .	4	0.105	-0.158	4.4283	0.351
. .	. .	5	0.073	0.069	4.5568	0.472
. * .	. ** .	6	-0.133	-0.222	5.0472	0.538
. ** .	. .	7	-0.23	-0.019	6.7658	0.454
. * .	. .	8	-0.113	-0.043	7.2628	0.509
. * .	. **** .	9	-0.153	-0.357	8.4027	0.494
. ** .	. ** .	10	-0.209	0.216	11.23	0.34
. .	. ** .	11	-0.065	-0.21	11.643	0.391
. .	. * .	12	0.031	0.081	11.834	0.459

表 3 的左边表明灰色预测模型的残差序列的自、偏相关系数变化，均在 2 倍标准差范围内，两者值都非常小。结合自相关系数判断法可知灰色预测模型残差序列为平稳白噪声。

表 3 的右边展示了相应的 P 值, 其值均大于 0.05。结合 P 值判断法可知灰色预测模型残差序列是一个白噪声序列, 相应信息已经充分提取。

4. 结束语

本文研究了灰色预测模型的基本建模步骤和残差序列的白噪声检验, 并通过 2020 年 2 月 6 日至 2020 年 2 月 19 日日本“钻石公主号”邮轮上新冠肺炎感染人数具体展示了如何实现相应的步骤。在本文中, 利用时间序列方法得到了模型残差为白噪声序列, 说明 GM(1, 1)模型将原始数据的信息完全提取了出来。

参考文献

- [1] Deng, J.L. (1982) Control Problems of Grey Systems. *Systems & Control Letters*, **1**, 288-294.
[https://doi.org/10.1016/s0167-6911\(82\)80025-x](https://doi.org/10.1016/s0167-6911(82)80025-x)
- [2] 王义闹. GM(1, 1)的直接建模方法及性质[J]. 系统工程理论与实践, 1988(1): 27-31.
- [3] 姜忠军. GM(1, 1)模型及其残差修正技术在土地承载力研究中的应用[J]. 系统工程理论与实践, 1995(5): 72-78.
- [4] 谭冠军. GM(1, 1)模型的背景值构造方法及应用(I) [J]. 系统工程理论与实践, 2000, 20(4): 98-103.
- [5] 穆勇. 无偏灰色 GM(1, 1)模型的直接建模法[J]. 系统工程与电子技术, 2003, 25(9): 1094-1095.
- [6] 王宇熹, 汪泓, 肖峻. 基于灰色 GM(1, 1)模型的上海城镇养老保险人口分布预测[J]. 系统工程理论与实践, 2010, 30(12): 2244-2253.
- [7] 卢捷, 李峰. 基于初始值和背景值改进的 GM(1, 1)模型优化与应用[J]. 运筹与管理, 2020, 29(9): 27-33.
- [8] 贾俊平, 何晓群, 金勇进. 统计学[M]. 第 7 版. 北京: 中国人民大学出版社, 2019.
- [9] 周永道, 王会琦, 吕王勇. 时间序列分析及应用[M]. 北京: 高等教育出版社, 2016.