

意大利新冠肺炎疫情预测及理性评估

戎翔 高彦梅 李宇 卫晓帆

目录

1	1 引言	2
1.1	1.1 背景及研究内容	2
1.2	1.2 文献综述	2
2	2 数据来源及描述统计	3
2.1	2.1 数据来源	3
2.2	2.2 意大利疫情基本描述	4
2.3	2.3 疫情下的意大利旅游经济	6
3	3 指数平滑法	7
3.1	3.1 指数平滑模型理论	7
3.2	3.2 指数平滑法的应用	7
4	4 基于 arima 模型对意大利新冠肺炎死亡人数的预测	7
4.1	4.1 方法介绍	7
4.2	4.2 ARIMA 模型的建立	7
4.3	4.3 统计分析	9
4.4	4.4 模型预测与应用	11
5	5 基于灰色系统 GM (1,1) 模型分析 2019-nCoV 对意大利旅游经济的影响	13
5.1	5.1 模型介绍	13
5.2	5.2 GM (1,1) 建模步骤:	13
5.3	5.3 模型求解	15
5.4	5.4 结果分析	17

1 1 引言	2
6 6 总结	18
7 参考文献	19

1 1 引言

1.1 1.1 背景及研究内容

2019 年 12 月,我国湖北省武汉市陆续出现多起不明原因的病毒性肺炎,一种新发现的冠状病毒随即被确认为导致武汉不明原因肺炎病例的病原体。2020 年 1 月 12 日,世界卫生组织将该病毒命名为“2019 新型冠状病毒 (2019-nCoV)”,同年 2 月 11 日,国际病毒分类委员会又将其正式更名为严重急性呼吸综合征冠状病毒 2(SARS-Cov-2),世界卫生组织于同日该病毒导致的疾病正式命名为 2019 新型冠状病毒疾病 (COVID-19),国家卫健委随后将新型冠状病毒肺炎的英文缩写标准化为 COVID-19。COVID-19 传染性极强,迅速从武汉蔓延至全国各地,并在境外多个国家和地区传播,波及全球 200 多个国家。

根据最新数据统计,全球新冠肺炎累计确诊病例数突破 731 万例。目前,全球有 8 个国家累计确诊数超过 20 万,其中,美国累计确诊病例已经突破 204 万例,意大利、巴西等 7 个国家的累计确诊病例最新突破 20 万。此次新冠肺炎疫情给全世界带来了前所未有的挑战,其后续影响难以估量。意大利作为世界旅游业较发达的国家之一,也不可避免地受到了严重的影响。这种影响不仅发生于意大利与其他国家之间,更发生于意大利国内的方方面面。因此,本文分别利用指数平滑以及 ARIMA 模型对意大利新冠肺炎情况进行了预测分析,并利用灰色系统分析方法具体定量分析新冠肺炎对 2020 年意大利旅游收入产生的损失。

1.2 1.2 文献综述

林硕蕾利用灰色系统理论建立 GM (1,1) 模型,由 1997-2002 年旅游外汇收入的平均值预测 2003 年旅游外汇收入平均值;并通过历史数据计算每个月的指标值与全年总值的关系,得出每个月旅游外汇收入所占全年旅游收入的权重,从而预测出正常情况下 2003 年每个月外汇旅游收入的指标值;最后将其与 2003 年每月外汇旅游收入的真实值进行比较,定量分析 SARS 对旅游外汇收入的影响 [1]。

王旭艳等人利用 1 月 23 日到 3 月 25 日湖北省疫情数据,采用平滑指数模型对湖北省新冠肺炎疫情变化趋势进行分析和预测。结果表明,采用指数平滑模型获得的拟合值与实际值的趋势基本吻合,模型拟合效果较好,预测结果表明在 4 月 2 日现存确诊病例数将减少至 1000 例以内,且主要为重症和危重症病例 [2]。

严婧,杨北方对湖北省丙肝疫情资料进行整理,将 2004 年 1 月—2014 年 12 月丙肝发病数据作为建模部分,2015 年 1—12 月发病数据作为模型验证分部,对丙肝不同年份发病率进行趋势卡方检验,建立 ARIMA 模型和指数平滑法,并进行模型的参数估计、拟合检验、计算和预测、检验与评价。将预测值与实际值进行比较,计算相对误差绝对值,相对误差绝对值越小说明预测越准确,然后进行型预测效果比较 [3]。

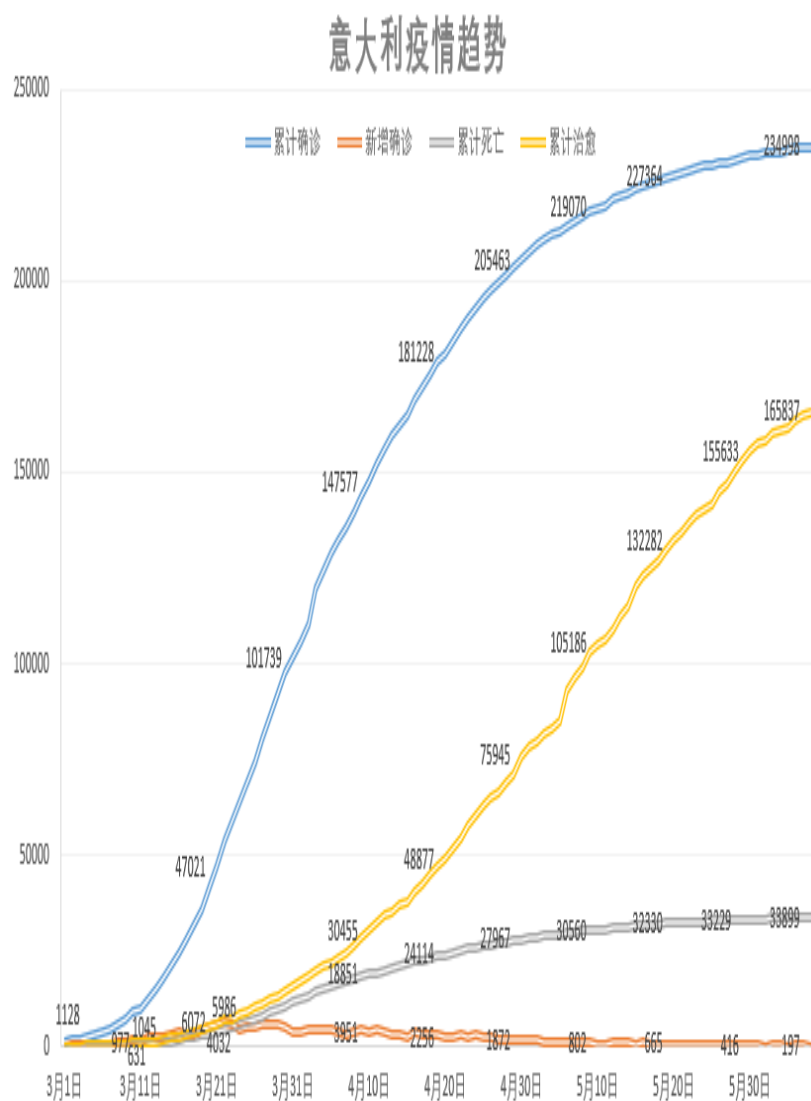
2 2 数据来源及描述统计

2.1 2.1 数据来源

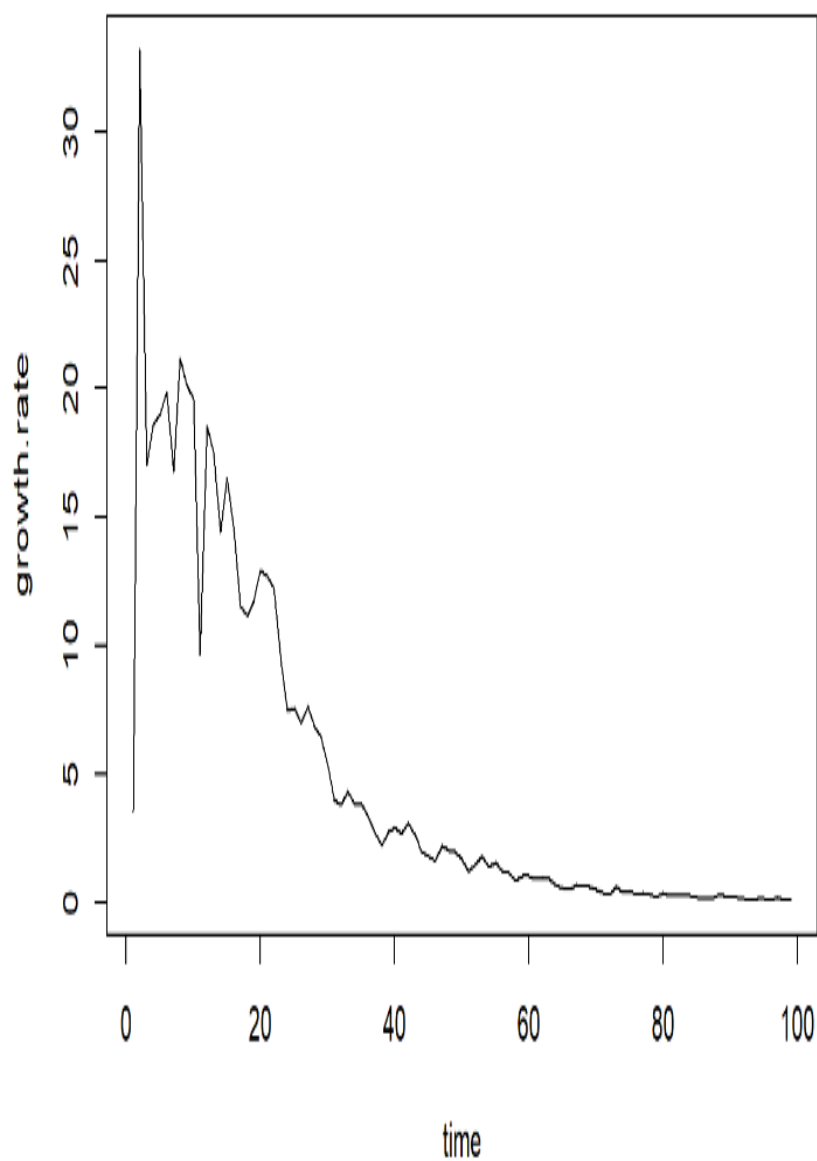
疫情数据来源于世界卫生组织官网 (<https://www.who.int/>),选取意大利 3 月 1 日-6 月 7 日的数据,具体包括累计确诊人数、新增确诊人数、累计治愈、新增治愈人数、累计死亡人数以及新增死亡人数。根据每日疫情数据,可以依据累计确诊人数、累计死亡人数以及累计治愈人数求得现存确诊人数,以及移除人数,其中现存确诊人数 = 累计确诊人数-累计死亡人数-累计治愈人数,移除人数 = 累计死亡人数 + 累计治愈人数。

有关意大利旅游收入的数据来源于 (<https://zh.tradingeconomics.com/>)。为了避免 2008 年金融危机对本次分析产生影响,我们选取 2009 年-2020 年的数据。

2.2 意大利疫情基本描述

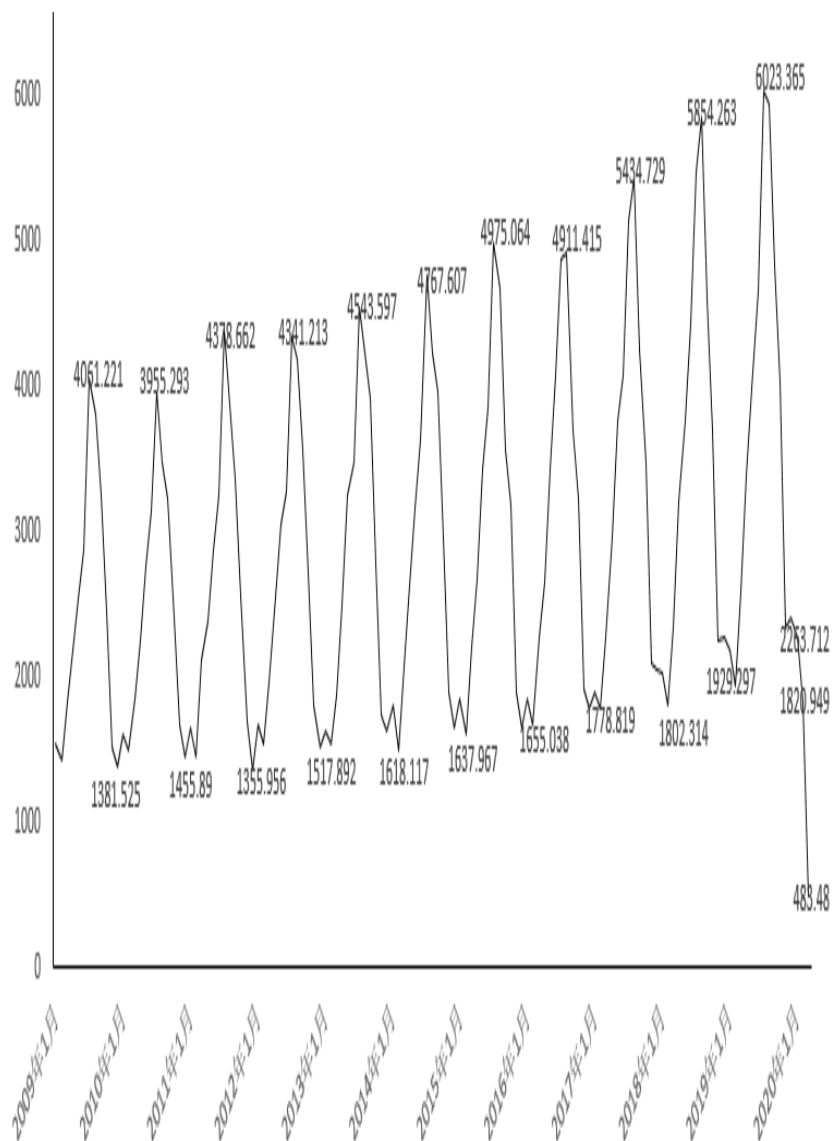


截至 6 月 7 日，意大利累计确诊人数达 23 万之多，累计病例数从 3 月初开始迅速增加，到 5 月份左右减缓，而后趋于稳定，尤其在近日以来，新增确诊病例数在 200 左右波动，并且在不断减少，说明意大利疫情得到了较好的控制。



随着新增确诊人数的先增后减，增长率曲线也呈现出先增后减的趋势，在 3 月 1 日-3 月 22 日期间，增长率大于 10%，而后逐渐减小，在 4 月 29 日以后增长率均小于 1%，并逐渐趋向于 0，说明意大利的 COVID-19 得到了有效控制。

2.3 疫情下的意大利旅游经济



意大利 2009-2020 年的旅游收入走势图显示，旅游收入逐年上升，并呈规律性的周期变化，在每年 7 月份左右旅游收入最高，1 月份左右最低。由于 COVID-19 的影响，意大利 2020 年的旅游收入在一月份下滑后并没有回升，甚至在 3 月份降到了 483.48 百万欧元，降到了近十年的最低点，可见 COVID-19 疫情对意大利经济有着不容忽视的影响。

3 3 指数平滑法

3.1 3.1 指数平滑模型理论

3.2 3.2 指数平滑法的应用

4 4 基于 arima 模型对意大利新冠肺炎死亡人数的预测

4.1 4.1 方法介绍

ARIMA 模型即自回归移动平均模型，时间序列分析模型通常是指求和自回归移动平均模型 (Autoregressive integrated moving average, ARIMA)，由统计学家 Box 和 Jenkins 于 1970 年首次系统提出，以数理统计和微积分等传统数学理论为基础建立起来的预测模型，是时间序列分析方法中应用比较广泛的方法之一，短期预测准确率较高。在病预测等方面得到广泛应用。根据序列特征，可分为平稳和非平稳时间序列。对于非平稳时间序列，可以通过 d 次差分运算将其转换为平稳时间序列，此时即为 ARIMA。ARIMA 模型的形式可表示为 $ARIMA(p,d,q)$ 。 p,d,q 分别表示自回归阶数、差分阶数和移动平均阶数。模型构建过程包括数据时间序列平稳化、模型识别、参数估计和模型检验。本研究收集今年 3 月 1 号到 6 月 7 号意大利疫情数据，用 ARIMA 模型对死亡人数进行模型构建、拟合，并应用最优模型参数预测意大利疫情发展趋势，以期发现意大利疫情的潜在变化规律，为开展防控工作提供理论依据。

4.2 4.2 ARIMA 模型的建立

4.2.1 (1) 平稳性和白噪声 (White noise) 检验

ARIMA 建模要求序列为平稳且是非随机性序列，若数据为非平稳时间序列，则可对原序列进行差分运算使其转换为平稳序列。白噪声检验，也称为纯随机性检验，主要用来检验数据是否为纯随机序列。本研究主要用单位根 ADF 的方法进行平稳性检验；计算 LB 统计量进行白噪声检验。

4.2.2 (2) 模型识别和定阶

对平稳非白噪声差分序列，依据自相关图（ACF）和偏自相关图（PACF）初步选择合适的模型参数进行拟合。模型定阶根据最小信息量原则（Akaike information criterion, AIC）和贝叶斯信息准则（Bayesian information criterion, BIC）进行，AIC 和 BIC 值最小的模型为最佳。

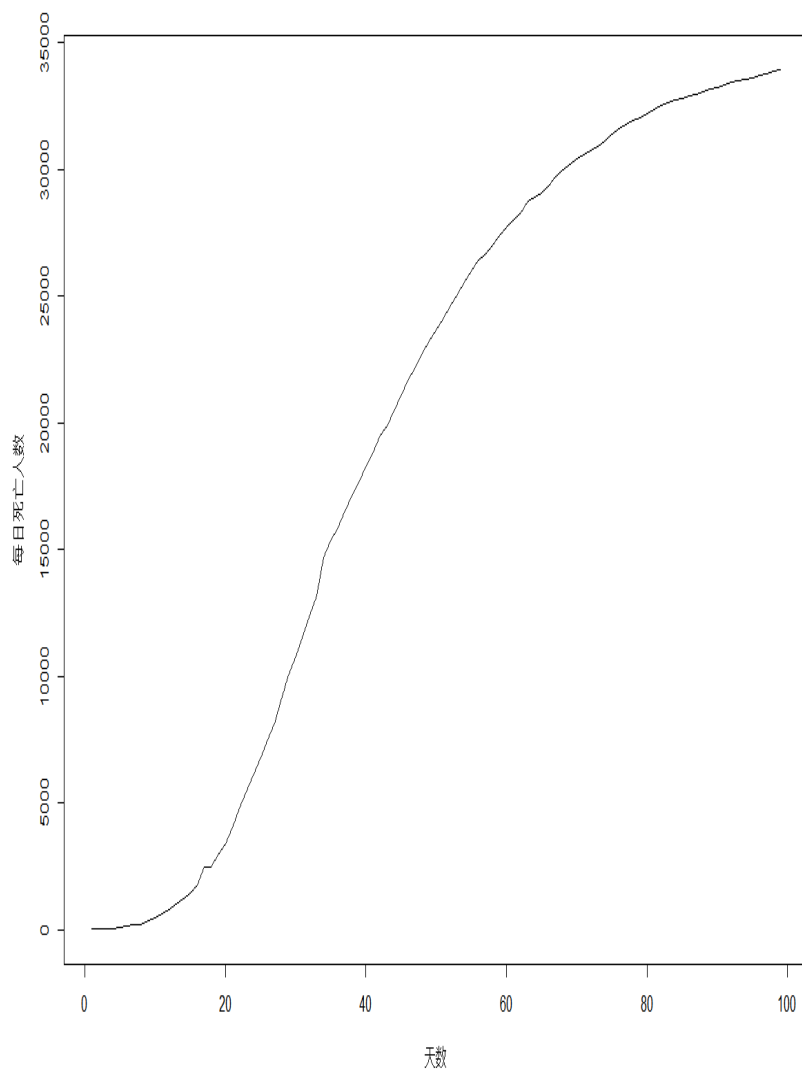
4.2.3 (3) 参数估计与模型检验

在模型识别阶段，采用最大似然法对粗模型进行参数估计与显著性检验。模型显著性检验即残差白噪声检验，是检验模型是否充分提取了序列信息，即残差是否为白噪声。若是，则构建的模型可以用于预测；否则，返回第 2 步重新。

4.2.4 (4) 模型预测

建立 3 月 1 号到 6 月 7 号意大利新冠肺炎死亡人数变化模型进行预测

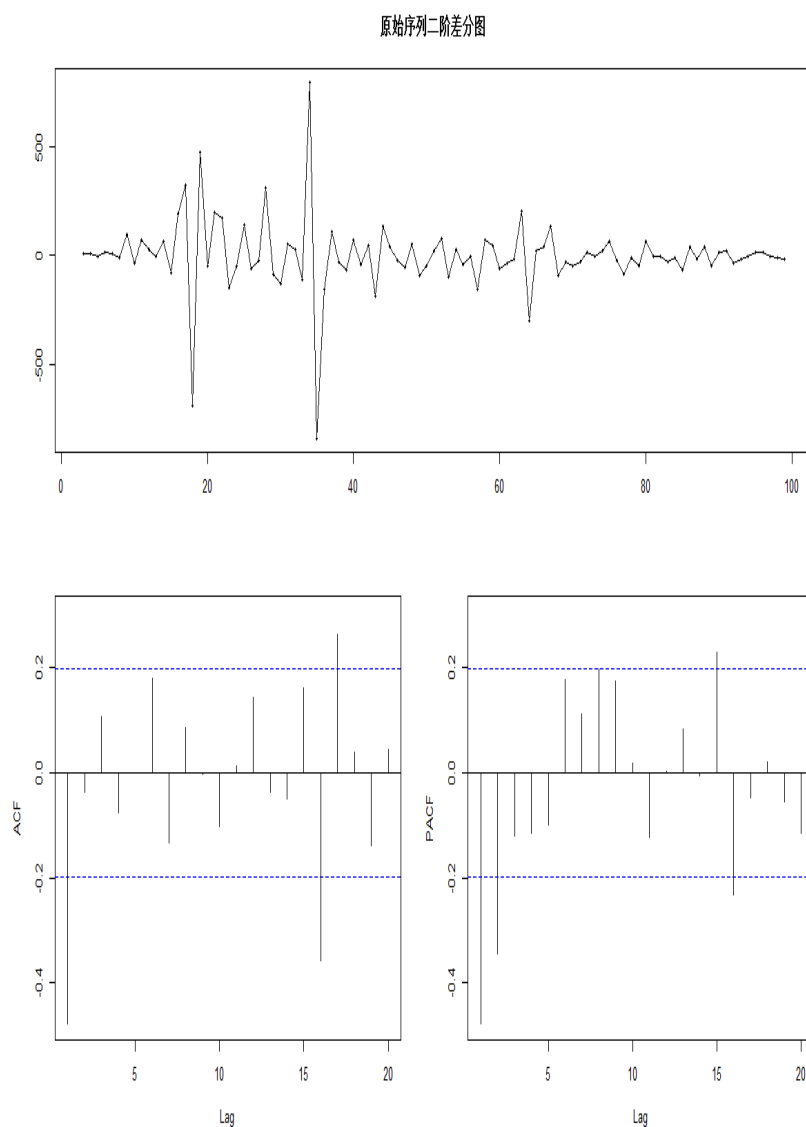
4.3 统计分析



意大利 3 月 1 号到 6 月 7 死亡人数数据绘制时间序列图, 可见明显的递增趋势, 长期而言呈递增速度在减慢, 为不平稳序列, 因此对其 2 阶差分以使其平稳.

尽管有文献提及肺炎发病具有季节性, 但就此次而言, 由于是新型肺炎, 之前未发生过, 数据也少, 故不宜进行季节性差分。根据 `ndiffs` 函

数可知原为二阶后平稳。经处理后，原始数据趋于平滑，对数据进行处理后的自相关函数（ACF）分析图与偏自相关函数（PACF）分析图，如下图所示。



经过 ADF 单位根检验，P 值小于 0.01，显示原始序列二阶差分后为平稳序列。经过白噪声检验，P 值小于 0.01，证明原始序列二阶差分后不是白噪声序列，即序列中包括有意义的数据，继续拟合模型。由 acf 图和

pacf 图以及 auto.arima 函数, 可尝试拟合 ARIMA(0,2,1)。ARIMA (2, 2, 0), ARIMA (1, 2, 2)

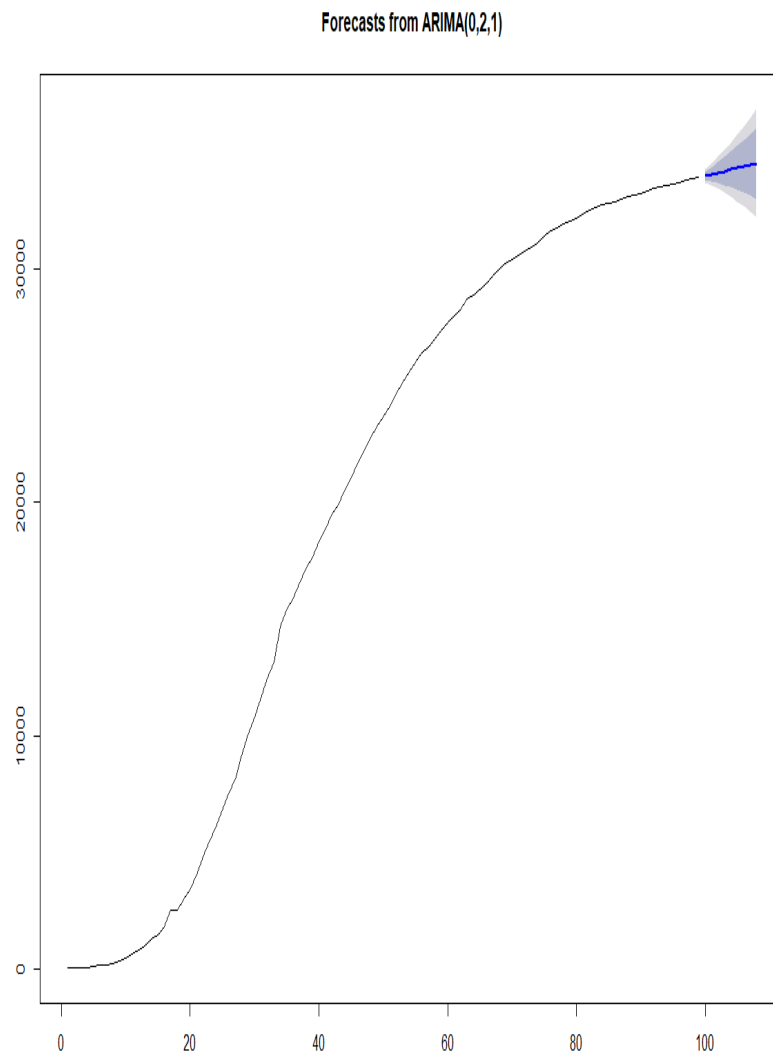
ARIMA (2, 2, 0) 的系数与标准误相除, 绝对值均大于 1.96, 系数是显著的, 残差通过白噪声检验, $aic = 1242.48$ $bic=1250.2$ 。

ARIMA (1, 2, 2) 虽然系数显著, 但是模型不可逆, 舍去. $AIC=1236.23$, $BIC=1246.53$

ARIMA (0, 2, 1) 系数显著. 残差也通过白噪声检验. $aic = 1240.92$, $bic=1246.069$ 经过比较, 最终选择 ARIMA(0,2,1)。

4.4 模型预测与应用

应用该预测模型对第 100 天 (6 月 8 号) 到第 108 天 (6 月 16 号) 进行预测, 其预测值分别为: 33966, 34034, 34101, 34169, 34236, 34303, 34371, 34438, 34506.



5 5 基于灰色系统 GM (1,1) 模型分析 2019-nCoV 对意大利旅游经济的影响

5.1 5.1 模型介绍

灰色模型 (GreyModel, 简称 GM), 即通过累加、累减等方式把原始数据变换成生成序列, 再利用其中的规律进行建模预测的方法, 可针对不同问题建立相应的灰色模型, 如 GM(1,1), GM(1,m), GM(n,m), 分别用于一个自变量一个因变量, 多个自变量一个因变量, 多个自变量多个因变量。模型的优点是可以利用较少的数据进行预测, 通过微分方程充分挖掘系统的本质, 精度较高, 可适用于中短期的预测。

其中, GM (1,1) 模型是灰色系统理论中的基本模型, 它将无规律的原始数据累加生成后, 使其变为有规律的生成数列, 然后建立相应的微分方程模型, 从而预测事物未来发展趋势和状态。

根据我们收集的历史数据可以看出, 在正常情况下, 全年的平均性能较好地反映出意大利旅游收入的变化规律, 据此把预测分成两部分:

(1) 利用灰色理论建立 GM (1,1) 模型, 由 2009- 2019 年的平均值预测 2020 年平均值;

(2) 通过历史数据计算每个月的旅游收入与全年总值的关系, 从而预测出正常情况下 2020 年每个月的旅游收入。

5.2 5.2 GM (1,1) 建模步骤:

5.2.1 数据的检验与处理

根据收集的数据计算出 2009-2019 年意大利旅游收入的平均值, 即模型的原始数据列, 记为 $x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$, 计算数列的级比:

$$\lambda = x^{(0)}(k-1)/x^{(0)}(k), k = 2, 3, \dots, n$$

若所有级比都落在可容覆盖区间 $x = (e^{-2/n+1}, e^{2/n+1})$ 内, 则数列可建立 GM (1,1) 模型。否则, 需要对数据做适当的变换处理, 如平移变换 $y^{(0)}(k) = x^{(0)}(k) + c, k = 1, 2, \dots, n$, 取 c 使得数据列的级比都落在可容覆盖区间 (0.846, 1.181) 内, 因此可以建立 GM (1,1) 模型。

5.2.2 累加生成

令 $x^{(1)}$ 为 $x^{(0)}$ 的一次累加序列, 则 $x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n))$, 其中, $x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i), k = 1, 2, \dots, n$.

5.2.3 均值生成

令 $z^{(1)}$ 为 $x^{(1)}$ 的紧邻均值生成序列, $z^{(1)} = (z^{(1)}(1), z^{(1)}(2), \dots, z^{(1)}(n))$, 其中, $z^{(1)}(k) = 1/2(x^{(1)}(k) + x^{(1)}(k+1)), k = 1, 2, \dots, n$.

5.2.4 GM (1,1) 模型

GM (1,1) 模型的灰微分方程为: $x^{(0)}(k) + az^{(1)}(k) = b$ 。式中, a 为发展系数, b 为灰作用量, 均为待定系数。

设待估参数向量 $\hat{a} = [\hat{a}, \hat{b}]^T$, 则灰微分方程的最小二乘估计参数列满足 $\hat{a} = [\hat{a}, \hat{b}]^T = (B^T B)^{-1} B^T Y$, 其中,

$$B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix}$$

$$Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}$$

则此模型的时间响应函数为 $x^{(1)}(t) = (x^{(1)}(1) - \frac{b}{a})e^{-at} + \frac{b}{a}$, 估计出参数 a 、 b 以后, 可以利用此函数对原始序列进行预测。

5.2.5 模型检验

在利用 GM (1,1) 模型进行预测前, 需要进行模型精度和误差的检验, 来保证预测的准确性。

5.2.5.1 残差检验

对估计值和实际值的残差进行逐点检验, 并计算原始序列 $x^{(0)}(i)$ 与 $\hat{x}^{(0)}(i)$ 的绝对残差序列 $\Delta^{(0)}(i) = |x^{(0)}(i) - \hat{x}^{(0)}(i)|$, 进而计算其相对残

差序列 $\phi_i = \frac{\Delta^{(0)}(i)}{x^{(0)}(i)}$, 最后得出整个模型的平均相对误差 $\bar{\phi} = \frac{1}{n} \sum_{i=1}^n \phi_i$, 将其与给定的残差水平进行比较, 若小于给定的残差水平, 则称模型为残差合格模型。

5.2.5.2 后验差检验

对残差分布的统计特性进行检验。首先, 分别计算原始序列的平均值 $\bar{X} = \frac{1}{n} \sum_{k=1}^n x^{(0)}(k)$ 和方差 $S_1 = \sqrt{\frac{1}{n} \sum_{k=1}^n [x^{(0)}(k) - \bar{X}]^2}$ 、残差的均值 $\bar{E} = \frac{1}{n-1} \sum_{k=2}^n E(k)$ 和方差 $S_2 = \sqrt{\frac{1}{n-1} \sum_{k=2}^n [E(k) - \bar{E}]^2}$ 、后验差比值 $C = \frac{S_2}{S_1}$ 、小概率误差 $P = P(|E(k) - \bar{E}| < 0.6745 S_1)$, 最后, 根据相应标准的临界值的大小, 综合评价模型的拟合效果, 常用的模型精度判断等级见表 1。

5.2.6 预测

根据历史数据计算出 2020 年每个月旅游收入占全年总值的比例 u_i , 即 $u_i = \sum_{j=1}^{11} a_{ij} / \sum_{i=1}^{12} \sum_{j=1}^{11} a_{ij}$, 在得到 2020 年的预测值后, 可算出 2020 年每个月旅游收入预测值。

5.3 模型求解

由 2009-2019 年意大利旅游收入的月度数据, 可计算得年平均数列及一次累加值数列:

$$X^{(0)} = (2404.70, 2388.09, 2574.22, 2671.31, 2757.81, 2853.37, 2962.99, 3029.93, 3262.90, 3476.03, 3691.85$$

$$X^{(1)} = (2404.70, 4842.79, 7417.01, 10088.32, 12846.14, 15699.51, 18662.49, 21692.43, 24955.33, 24831.35, 32$$

从而得到其紧邻均值生成序列:

$$Z^{(1)} = (3623.75, 6129.90, 8752.67, 11467.23, 14272.82, 17181.00, 20177.46, 23323.88, 266693.34, 30277.28, 32$$

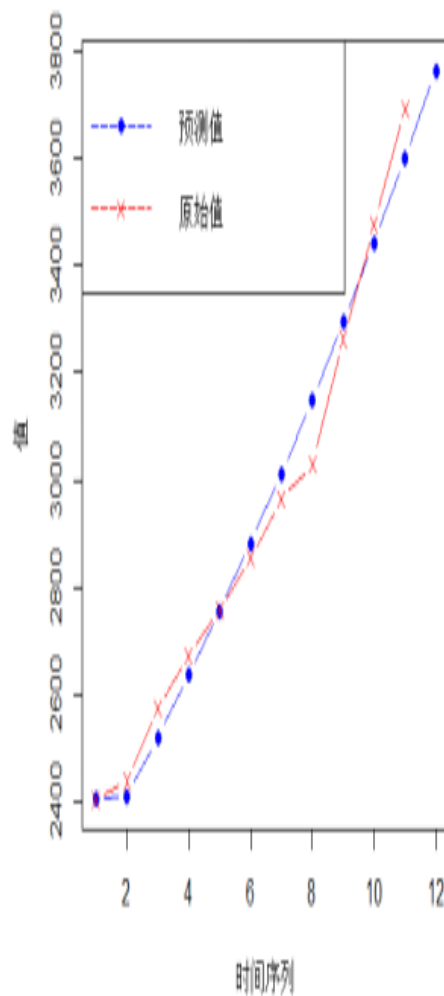
求得其发展系数和灰作用量: $a = -0.0445, b = 2251.087$, 则此模型的时间响应函数为

$$x^{(1)}(t) = 52990.92e^{0.0445t} - 50586.22$$

在利用上式进行预测之前, 对其模型精度和误差进行检验。在软件中计算得到, 残差检验中 $\phi = 0.0155$, 在 0.01 与 0.05 之间, 模型精度合

格；而在后验差检验中， $C = 0.0898$ ，小于 0.35，模型精度为优。因此，可以用此模型对 2020 年意大利月均旅游收入进行预测。

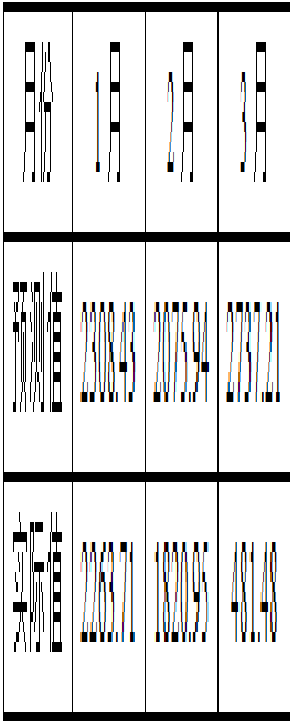
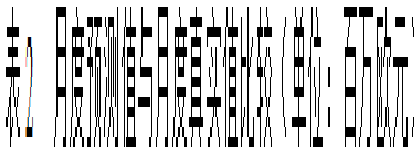
根据上式预测得到： $X^{(0)}(11) = 3763.723$ ，（即预测所得 2020 年全年旅游收入为 45164.68（百万）。预测结果如图 1 所示。



同时，2020 年第 i 个月的指标值占全年总值的比例为

$$u = (0.0511, 0.0459, 0.0606, 0.0759, 0.0934, 0.1056, 0.1363, 0.1313, 0.1095, 0.0868, 0.0539, 0.0496)$$

月度预测值与月度真实值比较结果见表 2。



5.4 5.4 结果分析

根据上表中的预测结果，意大利 1 月份的旅游收入应为 2308.43 百万欧元，但实际收入却为 2263.71 百万欧元，二者相差 44.72 百万欧元，差距可能并不明显；而预测意大利 2 月份的旅游收入应为 2075.94 百万欧

元, 实际收入为 1820.96 百万欧元, 差距达到了约 255 百万欧元; 对于三月份, 预测所得意大利旅游收入应为 2737.21 百万欧元, 但实际值却为 481.48 百万欧元, 差距扩大是非常明显的。由此可以说明, 此次疫情对意大利的旅游经济产生了很大的冲击, 这种冲击在三月份达到了最大。这可能是由于在 1 月份人们对新冠肺炎疫情不够了解所造成的, 在 1 月 23 日武汉封城后, 各国都采取了相应的防范措施, 因此对意大利旅游情况的冲击加大。

6 6 总结

针对意大利新冠肺炎疫情发展情况, 本文提出了指数平滑模型和 ARIMA 模型, 分别对意大利累计确诊人数、现存人数和死亡人数进行预测与分析, 并利用灰色模型 GM(1,1) 分析 COVID-19 疫情对意大利旅游经济的影响。

指数平滑模型结果显示, 6 月底, 意大利累计确诊人数可能会突破 24 万, 但总体来说, 累计确诊人数的增长变缓, 整体情况有明显改善。

模型对死亡人数的分析表明, 新冠肺炎死亡人数的增长率正日益交降低, 趋于平稳。从一开始的指数增长到平稳发展经历了一百天左右的过程, 说明意大利现行的疫情政策非常有效果。结合具体预测数据, 有理由相信意大利新冠疫情会在最近得到较为充分的控制。

此次疫情除了对意大利造成了严重的人员伤亡外, 也对当地的旅游经济产生了很大的冲击, 这种冲击在三月份达到了最大, 三月份旅游收入低落到近十年的最低点, 这对意大利旅游经济造成了巨大的冲击。但随着疫情的逐渐控制以及当地政府经济政策的实施, 相信意大利经济会有所好转。

综上所述, 意大利新冠肺炎疫情情况的改善与政策的实施有密不可分的关系, 早在 3 月 10 日, 意大利便开始实行封锁政策, 成为欧洲第一个封国的国家。随着全球疫情不断蔓延, 意大利疫情却开始呈现下降趋势, 可见, 在强有力的政策管控下, 意大利新冠肺炎疫情情况得到好转, 从最开始的累计确诊人数排名第二到目前的排名第七。因此, 在其他地区疫情较为严重且本国尚存在新增确诊病例的情况下, 防疫意识不可松懈, 仍应该保持警惕, 注重个人卫生, 在重振经济的同时继续做好防疫工作, 尽快取得真正的解放。

7 参考文献

- [1] 林硕蕾. 灰色系统在经济上的应用-定量分析 SARS 对旅游外汇收入的影响 [2] 王旭艳, 喻勇, 胡樱, 宇传华. 基于指数平滑模型的湖北省新冠肺炎疫情预测分析 [J]. 公共卫生与预防医学, 2020, 31(01): 1-4. [3] 严婧, 杨北方. 指数平滑法与 ARIMA 模型在湖北省丙型病毒性肝炎发病预测中的应用 [4] 张文斗, 祖正虎, 许晴, 徐致靖, 郑涛. SARS 疫情对中国交通运输业和电信业的影响分析 [J]. 军事医学, 2012, 36(10): 762-764+800. [5] 林硕蕾, 王凌冉. 灰色系统在经济上的应用——定量分析 SARS 对旅游外汇收入的影响 [J]. 北京工业职业技术学院学报, 2005(04): 81-83+87. [6] 张世强, 张飒岚, 蒋峥, 王春丽, 王婷. 基于信息挖掘的麻疹流行趋势 GM(1,1) 评价模型 [C]. 中国电子学会信息论分会. Proceedings of 2010 Asia-Pacific Conference on Information Theory (APCIT 2010). 中国电子学会信息论分会: 中国电子学会信息论分会, 2010: 335-338. [7] 金启轩. 中国新冠肺炎疫情预测建模与理性评估 [J]. 统计与决策, 2020, 36(05): 11-14.