

2020 年南宁市流行性腮腺炎 ARIMA 模型预测研究

唐金芳¹, 王佑新², 农初师¹, 甘文烨¹, 曾小云²

1. 南宁市疾病预防控制中心, 广西 南宁 530028; 2. 广西医科大学研究生学院, 广西 南宁 530021

摘要: **目的** 建立自回归移动平均模型(auto-regressive integrated moving average, ARIMA)预测南宁市流行性腮腺炎的发病率,为防控流行性腮腺炎提供理论依据。 **方法** 采用 SPSS 23.0 软件,使用南宁市 2013 年 1 月—2019 年 6 月流行性腮腺炎月发病率建立 ARIMA 模型,运用 2019 年 7—12 月的实际值与预测值进行评估,并预测 2020 年 1—6 月的发病率。 **结果** 拟合的 ARIMA(1,0,0)(0,1,1)₁₂ 模型最佳,MAE = 0.695, RMSE = 1.230, MAPE = 23.487, 经检验残差序列为白噪声,拟合模型对短期有良好的预测效果。2020 年 1—6 月的预测月发病率分别为 9.88/10 万、7.62/10 万、8.32/10 万、9.40/10 万、11.91/10 万、11.67/10 万。 **结论** ARIMA 模型可用于南宁市流行性腮腺炎发病趋势的短期预测,适宜推广使用。

关键词: 流行性腮腺炎; ARIMA 模型; 预测

中图分类号: R512.1 **文献标识码:** A **文章编号:** 1006-3110(2021)03-0313-04 **DOI:** 10.3969/j.issn.1006-3110.2021.01.017

Application of ARIMA model to predicting the incidence rate of mumps in Nanning City, 2020

TANG Jin-fang¹, WANG You-xin², NONG Chu-shi¹, GAN Wen-ye¹, ZENG Xiao-yun²

1. Nanning Municipal Center for Disease Control and Prevention, Nanning, Guangxi 530028, China;

2. Graduate School of Guangxi Medical University, Nanning, Guangxi 530021, China

Corresponding author: ZENG Xiao-yun, E-mail: zxyxjw@21cn.com

Abstract: **Objective** To establish an auto-regressive integrated moving average (ARIMA) for predicting the incidence rate of mumps in Nanning City, and to provide a theoretical basis for mumps prevention and control. **Methods** SPSS 23.0 software was used to establish the ARIMA model based on the monthly incidence rates of mumps in Nanning City from January 2013 to June 2019. The actual and predicted values from July to December in 2019 were applied to assessing the model, and the incidence rates of mumps from January to June in 2020 were predicted. **Results** The fitted ARIMA (1,0,0) (0,1,1)₁₂ model was the best, with mean absolute error being 0.695, root mean squared error being 1.230, and mean absolute percentage error being 23.487. The residual sequence was classified as white noise, and the fitted model had good short-term prediction effect. The predicted monthly incidence rates of mumps from January to June in 2020 were 9.88/100,000, 7.62/100,000, 8.32/100,000, 9.40/100,000, 11.91/100,000 and 11.67/100,000, respectively. **Conclusions** The ARIMA model can be used for short-term prediction of epidemic trend of mumps in Nanning City, and it is suitable for popularization and application.

Keywords: mumps; ARIMA model; prediction

流行性腮腺炎是一种由腮腺炎病毒引起的急性呼吸道传染病^[1],儿童和青少年多发,临床表现主要为非化脓性腮腺肿痛、发热等症状^[2],病毒可侵犯消化系统、神经系统和生殖系统引起胰腺炎、脑膜炎、睾丸炎等多种并发症。流行性腮腺炎主要是通过飞沫传播,传染性强,极易在学校引起暴发流行。因此,探讨流行性腮腺炎的流行趋势及疾病预测,对于流行性腮腺炎防控策略及措施的制定具有十分重要的现实意义。

基金项目:广西壮族自治区卫生健康委员会科研课题(课题编号:Z20201482)

作者简介:唐金芳(1978-),女,湖南武冈人,硕士,主任医师,研究方向:疾病控制。

通信作者:曾小云, E-mail: zxyxjw@21cn.com。

义。目前,常用的方法之一就是采用 ARIMA 模型预测传染性疾病,这种方法将时间序列视为一组随着时间变化的变量,可以通过时间序列的历史值来预测其未来值^[3]。南宁市是广西流行性腮腺炎的高发地区之一^[4],一直以来未有针对该地区的模型预测等方面的研究。本文根据 2013 年 1 月—2019 年 6 月南宁市流行性腮腺炎月发病率序列有线性趋势的特点,建立 ARIMA 模型,对南宁市流行性腮腺炎月发病率做短期预测,实现动态监测流行性腮腺炎的发病趋势,建立预警机制,为制定防控策略提供新思路。

1 材料与方法

1.1 数据来源 南宁市流行性腮腺炎发病数据来自

中国疾病监测信息报告系统,人口数据来源于南宁市统计局。

1.2 统计学方法 运用 Excel 2010 整理分析数据,并绘制统计图;用 SPSS 23.0、Eviews 8.0 建立 ARIMA 模型并进行预测。

1.3 ARIMA 模型构建预测

1.3.1 ARIMA 模型简介 ARIMA 模型记为 $ARIMA(p, d, q)(P, D, Q)_s^{[5]}$ 。在该模型中,AR 是“自回归”项,MA 为“移动平均”项, p, P 分别表示自回归和季节性自回归的阶数, d, D 分别表示差分和季节性差分阶数, q, Q 表示移动平均和季节性移动平均阶数, s 代表季节性周期。

1.3.2 建模步骤

1.3.2.1 平稳性判定与平稳处理 观察时间序列图的趋势特征,并进行单位根检验,常用增项 DF 单位根检验(augmented Dickey-Fuller test, ADF),即 ADF 检验。序列不平稳,可进行 d 次差分转化为平稳序列。

1.3.2.2 模型识别与定阶 常根据 ARIMA 模型自相关函数(autocorrelation function, ACF)和偏自相关函数(partial autocorrelation function, PACF)的特征,选择合适的模型形式和滞后阶数进行参数估计,利用贝叶斯信息准则(Bayesian information criteria, BIC)进行定阶。

1.3.2.3 模型参数估计和检验 常应用最小二乘法估计或最大似然估计。检验残差序列 ACF 和 PACF 是否超出 95% 的可信限。并检验残差是否为白噪声,经检验,Box-Ljung Q 统计量没有显著性,说明建立的模型已经将时间序列中的相关信息提取完全,可用于预测。

1.3.3 模型预测评价 用平均绝对误差(MAE)、均方根误差(RMSE)、相对百分误差绝对值的平均值(MAPE)来评估模型的拟合和预测的效果。

2 结果

2.1 流行和分布基本概述 2013—2019 年南宁市流行性腮腺炎共报告病例数为 21 941 例,无死亡病例报告。各年度报告发病数(发病率)分别为 2 986 例(43.2/10 万)、1 531 例(22.41/10 万)、1 652 例(24.01/10 万)、1 939 例(27.76/10 万)、3 173 例(43.93/10 万)、3 674 例(52.02/10 万)和 6 986 例(96.3/10 万),除 2013 年外,发病呈逐年上升趋势。暴发疫情 45 起,发病 2 458 人,其中小学 26 起,中学 19 起。

2013—2016 年发病时间分布趋势整体相近,

2017—2018 年发病趋势上升明显,月发病高峰和低谷明显,季节性特点突出,2019 年发病趋势上升更为明显,月发病高峰和低谷更为明显,季节性特点更为突出。整体上,年度发病呈两个波动高峰,两个发病低谷,第一个高峰出现在每年的 4—7 月,第二个高峰出现在每年的 10—12 月,两个发病低谷分别为每年的 2 月和 8—9 月,见图 1。发病人群年龄分布有较大的差异,所有年龄人群均可发病,但发病主要集中在儿童及青少年,其中,0~14 岁人群占总发病人群 84.60%,学生、幼托儿童、散居儿童占总报告数 90.50%。

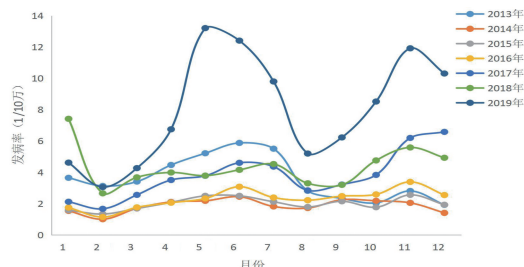


图 1 2013—2019 年南宁市流行性腮腺炎发病的时间分布

2.2 拟合 ARIMA 模型

2.2.1 平稳性判断 用 SPSS 23.0 对 2013 年 1 月—2019 年 6 月的南宁市流行性腮腺炎月发病率作序列图,见图 2。可发现随着时间的变化,其发病率有一定的向上波动性和周期性变化。首先,对其随时间波动的发病率进行单位根检验,ADF 检验结果显示, $t = -3.9687, P = 0.0158 < 0.05$,拒绝原假设,即序列不存在单位根,可认为原序列已经满足平稳性要求,但系列仍然受到一定的季节周期性影响,进行一阶季节性差分,消除季节性影响,使序列更满足平稳性要求,见图 3。

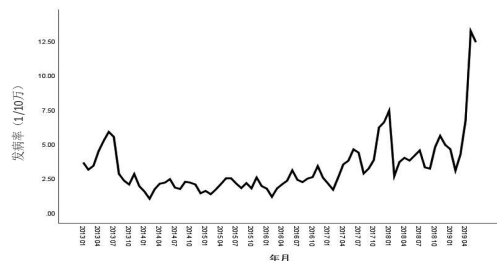


图 2 2013—2019 年南宁市流行性腮腺炎月发病率序列图

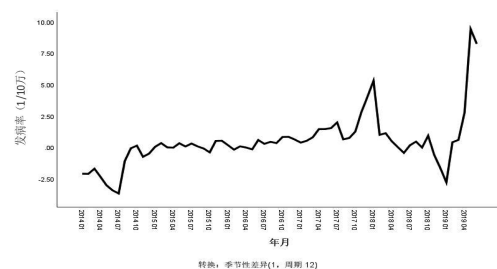


图 3 2013—2019 年南宁市流行性腮腺炎一阶季节差分月发病率序列图

2.2.2 模型识别与定阶 对经过一阶周期性转换的序列绘制 ACF 图与 PACF 图,见图 4、图 5。ACF 图显示 ACF 呈明显拖尾状态;PACF 图显示一阶以后 PACF 趋向于零,PACF 截尾。初步确定 ARIMA 模型参数表达式为 $ARIMA(1,0,0)(P,1,Q)$ 。有关文献显示^[6-7],P、Q 阶数一般不超过 2,可以采用逐个代入尝试的方法,见表 1。根据标准化 BIC,标化 BIC 值越小,Ljung-Box Q 检验结果无统计学意义($P>0.05$)^[8],说明模型的拟合优度越好,综合以上各指标最终确定 $ARIMA(1,0,0)(0,1,1)_{12}$ 为最佳模型。

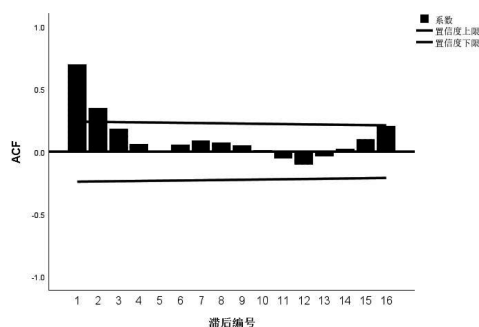


图 4 南宁市流行性腮腺炎一阶季节差分后的 ACF 图

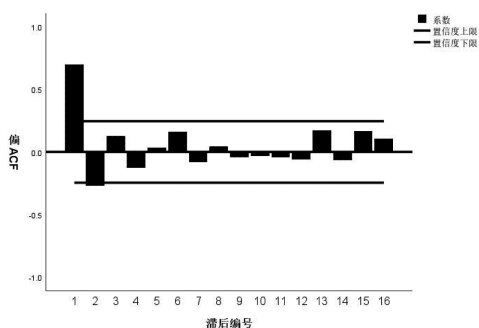


图 5 南宁市流行性腮腺炎一阶季节差分后的 PACF 图

表 1 列举备选模型的标化 BIC、平稳 R 方及 Ljung-Box Q 检验结果

备选模型	标准化 BIC 值	Ljung-Box Q 检验		
		平稳 R 方	统计量	P 值
$ARIMA(1,0,0)(0,1,0)_{12}$	0.696	0.613	14.136	0.657
$ARIMA(1,0,0)(1,1,0)_{12}$	0.623	0.668	12.273	0.725
$ARIMA(1,0,0)(1,1,1)_{12}$	0.699	0.669	12.351	0.652
$ARIMA(1,0,0)(0,1,1)_{12}$	0.605	0.674	12.232	0.728
$ARIMA(1,0,0)(0,1,2)_{12}$	0.692	0.671	12.142	0.668
$ARIMA(1,0,0)(2,1,0)_{12}$	0.702	0.668	12.305	0.656
$ARIMA(1,0,0)(2,1,1)_{12}$	0.782	0.668	12.319	0.581
$ARIMA(1,0,0)(1,1,2)_{12}$	0.760	0.675	9.505	0.545
$ARIMA(1,0,0)(2,1,2)_{12}$	0.833	0.677	12.444	0.492

2.2.3 参数估计和假设检验 $ARIMA(1,0,0)(0,1,1)_{12}$ 模型的常数项,自回归项均有统计学意义,见表 2。残差序列相关图上 ACF 和 PACF 不超出 95% 的可信

限,说明残差序列为白噪声,其 Box-Ljung Q 统计量为 12.232, $P=0.728>0.05$,提示没有显著性,说明建立的模型合适,可用于预测。

表 2 南宁市流行性腮腺炎 $ARIMA(1,0,0)(0,1,1)_{12}$ 模型参数

参数	估计	SE	t 值	Sig
常数	0.595	0.269	2.211	0.032
AR 滞后 1	0.680	0.110	6.160	0.000
AR, 季节性滞后 1	1.000	-	-	-

2.2.4 ARIMA 模型的结果显示 运用 $ARIMA(1,0,0)(0,1,1)_{12}$ 预测南宁市 2019 年下半年流行性腮腺炎的发病率,预测值与实际值的比较,见表 3。流行性腮腺炎发病率模型拟合趋势与实际发病趋势整体基本一致,除 8 月份流行性腮腺炎发病率的降幅较大外,实际发病均在预测发病的 95% 置信区间内,预测效果可靠。进一步预测 2020 年 1—6 月的流行性腮腺炎发病率,仍将呈高发病水平,见图 6。

表 3 模型 $ARIMA(1,0,0)(0,1,1)_{12}$ 预测分析 (1/10 万)

2019 年	预测值	实际值	误差	95% 置信区间
7 月	11.77	9.20	2.57	9.44~14.11
8 月	10.21	5.20	5.01	7.05~13.37
9 月	9.88	6.28	3.60	6.16~13.60
10 月	10.23	8.52	1.71	6.10~14.36
11 月	10.99	11.92	-0.93	6.54~15.44
12 月	10.20	10.31	-0.11	5.51~14.90

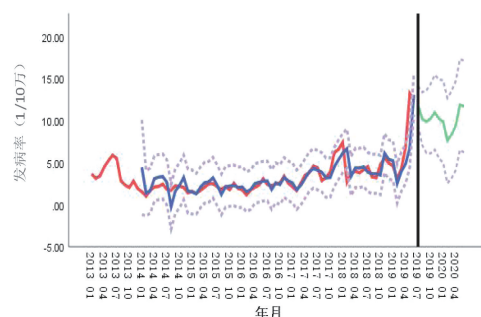


图 6 2020 年上半年南宁市流行性腮腺炎发病率模型拟合趋势图

3 讨论

2013—2019 年南宁市流行性腮腺炎年均发病率为 41.76/10 万,高于广西平均水平,和桂林市相近^[9],但低于 2005—2011 年南宁市流行性腮腺炎年平均报告发病率为 66.32/10 万^[10]。南宁市是广西壮族自治区首府,是人口最多交通流量最大的城市,人口流动性大,极易造成流行性腮腺炎等呼吸道传染病的暴发流行。从 2008 年开始,广西将麻疹风疹腮腺炎三联疫苗纳入国家免疫规划,18~24 月龄的儿童免费接种 1 剂麻疹风疹腮腺炎三联疫苗,南宁市流行性腮腺炎预控

成效还是极为显著。但从 2017 年开始,流行性腮腺炎发病大幅上升。为此,建立有效的预测模型,对流行性腮腺炎进行动态分析和预测显得尤为重要。

近年来,学者们运用神经网络、指数平滑法,logistic 微分方程模型,ARIMA 模型或组合模型等多种数学模型来进行流行性腮腺炎发病预测^[11-12]。其中应用最广泛、使用最多的是 ARIMA 模型^[13]。该模型方法利用事物的发展具有一定的惯性变化趋势,基于不同时期的数据之间存在相关性而建立预测模型,综合考虑了序列的趋势变化、季节周期性变化和白噪声因素等影响,并借助模型参数进行量化,经过重复判别修正,确定预测效果最佳的模型。同时,该模型还综合考虑了回归分析和移动平均的特点,是一种预测精度较高的中短期预测方法^[14]。本研究以 2013 年 1 月—2019 年 6 月流行性腮腺炎的月发病率数值共 78 个拟合 ARIMA 模型,经平稳 R 方、BIC 值、Ljung-Box Q 检验等综合考虑下获得了较佳的 ARIMA(1,0,0)(0,1,1)₁₂ 模型,MAE = 0.695, RMSE = 1.230, MAPE = 23.487。运用 ARIMA(1,0,0)(0,1,1)₁₂ 预测 2019 年下半年南宁市流行性腮腺炎的发病率,流行性腮腺炎发病率模型拟合趋势与实际发病趋势整体基本一致,除 8 月份流行性腮腺炎发病率的降幅较大外,流行性腮腺炎实际月发病率均落在模型预测值 95% 可信限内,预测效果可靠。显示该模型在南宁市流行性腮腺炎发病率的预测上效果良好,可用于该疾病的预测。在原模型的基础上,进一步预测 2020 年 1—6 月的流行性腮腺炎发病率,结果显示,2020 年 1—6 月的预测月发病率分别为 9.88/10 万、7.62/10 万、8.32/10 万、9.40/10 万、11.91/10 万、11.67/10 万。这提示,2020 年上半年南宁市流行性腮腺炎的发病仍呈处于较高流行水平,当实际月发病率超出预测值的 95% 可信区间的上限时,流行性腮腺炎疫情有全面暴发流行的风险^[15],应及时做好疫情研判,迅速采取应急防控措施,防止疫情进一步扩大。

基于本研究的结果,预测南宁市 2020 年上半年流行性腮腺炎发病仍处于高发态势,建议各县区应加强对流行性腮腺炎疫情监测和预警预测,针对性地开展防控工作,如在重点人群的中小学校广泛开展防治知识宣教工作,做好病人隔离,落实晨检和严格复课制定等。发生流行性腮腺炎暴发疫情时,及时对疫点和周边学校的易感人群开展应急接种。除做好麻疹风疹腮腺炎三联疫苗常规免疫和查漏补种工作外,还可

参考北京、上海等地^[16-17],选择单苗或含腮腺炎疫苗成分的联合疫苗开展补充免疫工作,对已接种 1 针麻疹风疹腮腺炎三联疫苗的适龄儿童,可及时补种第 2 针,消除免疫空白,筑牢免疫屏障。

参考文献

- [1] 李兰娟,任红. 传染病学[M]. 第 8 版. 北京:人民卫生出版社, 2016:73-77.
- [2] Saha I, Haldar D, Paul B, et al. An epidemiological investigation of mumps outbreak in a slum of Kolkata[J]. J Commun Dis, 2012, 44 (1):29-36.
- [3] 徐国祥. 统计预测和决策[M]. 上海:上海财经大学出版社, 1998:150-177.
- [4] 黎沙,杜进发,钟革,等. 2005—2015 年广西流行性腮腺炎流行病学特征分析[J]. 应用预防医学,2017,23(5):392-394.
- [5] 彭月. ARIMA 模型的介绍[J]. 电子世界,2014(10):259.
- [6] 宋媛媛,王雷,熊甜,等. ARIMA 模型与 GM(1,1)模型在痢疾发病数预测中的比较研究[J]. 实用预防医学,2019,26(7):888-892.
- [7] 张晓峰,尤爱国,潘静静,等. ARIMA 模型在三门峡手足口病发病预测中的应用[J]. 实用预防医学,2020,27(2):168-170.
- [8] 刘桂芬. 医学统计学[M]. 第 2 版. 北京:中国协和医科大学出版社,2007:354-360.
- [9] 蒋芹,喻国旗,覃健. 2005—2016 年桂林市流行性腮腺炎的流行特征和时空分析[J]. 现代预防医学,2019,46(4):588-592.
- [10] 汤洪洋,闭志友,林健燕,等. 2005—2011 年南宁市流行性腮腺炎流行特征[J]. 职业与健康,2013,29(2):208-209.
- [11] 曹磊,张义,刘峰,等. ARIMA-MLP 与 ARIMA-RBF 模型在流行性腮腺炎发病预测中的应用[J]. 公共卫生与预防医学, 2016,27(2):26-30.
- [12] 谢知,陈田木,林鑫,等. 流行性腮腺炎 logistic 微分方程预警模型的建立与应用[J]. 公共卫生与预防医学,2015,26(5):14-17.
- [13] 庞艳蕾,张惠兰,李向云,等. 灰色模型 GM(1,1)和 ARIMA 在拟合全国婴儿、5 岁以下儿童死亡率中的应用[J]. 中国卫生统计,2015,32(3):461-463.
- [14] 晁灵,冯晖,桂立辉. 2009—2013 年焦作市流行性腮腺炎流行特征分析及 ARIMA 预测[J]. 现代预防医学,2015,42(6):966-969.
- [15] 陈欣,吴晓敏,包名家,等. 佳木斯市 2004—2017 年流行性腮腺炎流行特征分析及趋势预测[J]. 中华疾病控制杂志,2019, 23(2):185-190.
- [16] 张丽文,朱宗龙,苑新海,等. 北京市昌平区 6 岁儿童加强接种前后麻疹、流行性腮腺炎和风疹抗体水平监测[J]. 疾病监测, 2018,33(1):63-65.
- [17] 陆红梅,沈金花,陈文花,等. 上海市松江区健康人群麻疹、风疹、流行性腮腺炎抗体水平监测[J]. 中国生物制品学杂志, 2016,29(1):51-56.

收稿日期:2020-04-01