

# 武汉市江汉区细菌性痢疾发病率的 ARIMA 模型预测应用

冯冰, 张艳, 周方<sup>△</sup>

(武汉市江汉区疾病预防控制中心, 武汉 430015)

**[摘要]** **目的** 探讨 ARIMA 模型在武汉市江汉区细菌性痢疾月发病率预测应用的可行性, 为该区细菌性痢疾的防控提供科学参考依据。**方法** 使用 SAS9.2 软件对 2005~2013 年该区菌痢月发病资料拟合 ARIMA 模型, 利用建立的模型对 2014 年 1~6 月菌痢月发病率资料进行预测和效果评价。**结果** 建立 ARIMA(0,1,1)X(0,1,1)<sub>12</sub> 模型拟合效果较好, 预测 2014 年上半年疫情将呈缓慢上升趋势, 预测值与实际值拟合趋势基本一致。**结论** ARIMA 模型可以作为该区菌痢月发病水平的短期预测模型。

**[关键词]** 细菌性痢疾; ARIMA 模型; 预测

**[中图分类号]** R516.4

**[文献标识码]** A

**[文章编号]** 1671-8348(2015)29-4123-02

## Application of ARIMA model to forecasting bacillary dysentery in Jianghan District of Wuhan City

Feng Bing, Zhang Yan, Zhou Fang<sup>△</sup>

(Center of Disease Control and Prevention of Jianghan District, Wuhan, Hubei 430015, China)

**[Abstract]** **Objective** To explore the feasibility of the application of autoregressive integrated moving average (ARIMA) model to forecast the monthly incidence rate of bacillary dysentery in Jianghan District of Wuhan, and to provide scientific references for prevention and control. **Methods** SAS 9.2 software was used to fit the ARIMA model for the monthly incidence rate of bacillary dysentery from 2005 to 2013. Thereafter, the fitted model was used to forecast the data from January to June in 2014, and was evaluated by comparing the actual incidence. **Results** The ARIMA(0,1,1)X(0,1,1)<sub>12</sub> model had a better fitted effect, it was predicted that the incidence of bacillary dysentery of Jianghan District of Wuhan City from January to June 2014 would increase slightly, and the forecast data of bacillary dysentery from January 2005 to December 2013 consisted with the actual situation. **Conclusion** ARIMA model can be used as short-term forecasting model to predict the monthly incidence rate of bacillary dysentery in Jianghan District of Wuhan.

**[Key words]** bacillary dysentery; ARIMA; forecasting

细菌性痢疾(菌痢)是一种由痢疾杆菌引起的常见消化系统传染病(乙类), 人群普遍易感, 尤其儿童居多<sup>[1]</sup>。江汉区是武汉市 7 个中心城区中人口密度最大的地区, 菌痢发病率一直占据法定报告传染病的前 5 位, 被细菌污染的食物和水是造成突发公共卫生事件的潜在威胁。研究发现监测资料的分析与预警是预防控制并减少传染病危害的关键举措<sup>[2-5]</sup>, 而时间序列分析 ARIMA 模型作为传染病发病预测的重要工具已得到广泛应用<sup>[6-9]</sup>, 为此本文利用江汉区菌痢发病资料建立时间序列模型, 探讨该模型预测菌痢发病率的可行性, 为该区菌痢的防控提供科学依据。

### 1 资料与方法

**1.1 一般资料** 2005~2013 年江汉区细菌性痢疾月发病率数据来自中国疾病预防控制中心信息系统, 2005~2013 年江汉区常住人口数来自武汉市统计局发布的人口统计年鉴。

**1.2 方法** 以 2005 年 1 月至 2013 年 12 月的发病率数据建立模型, 用 2014 年 1~6 月发病率数据用于检验模型的预测效果。

**1.3 统计学处理** 应用 SAS9.2 软件进行 ARIMA 建模; 主要有序列的平稳性检验、平稳化处理; 平稳序列的白噪声检验拟合 ARIMA 模型; 模型诊断、评价; 预测检验, 以  $P < 0.05$  为差异有统计学意义。

### 2 结果

**2.1 序列平稳化** 根据月发病率数据和人口统计资料计算得到 2005~2014 年菌痢月发病率数据, 菌痢发病每年都有明显

的季节趋势, 每年 6~8 月为发病高峰期, 2011~2013 年发病率较之前有下降的趋势。故对原始时间序列进行一阶差分 and 一次季节性差分后, 消除季节性的趋势影响, 使得序列平稳。

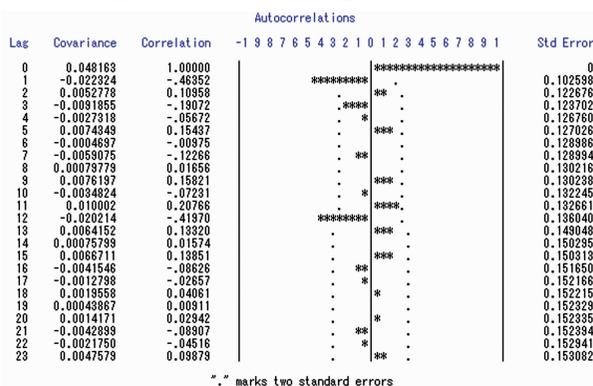


图 1 1 阶 12 步差分后自相关图

**2.2 模型识别与定阶** 根据差分后的时间序列自相关(ACF)图的性质(图 1), 样本自相关系数滞后 12 阶显著在 2 倍标准差之外, 说明差分后仍有显著的季节效应, 在滞后 1 阶自相关和偏自相关系数值都在 2 倍标准差之外, 说明序列还具有短期相关性, 当滞后阶数大于 1 时, 自相关系数出现了骤减, 显示 1 阶截尾, 偏自相关系数基本都在 95%CI 内, 呈现一定的衰减。因此, 根据 Box-Jenkins 理论<sup>[6,10-11]</sup>, 初步尝试拟合乘积季节模型 SARIMA(0,1,1)X(0,1,1)<sub>12</sub>, SARIMA(1,1,1)X(1,1,1)

表 1 备选 SARIMA 模型检验结果

参数	SARIMA(0,1,1)X(0,1,1) <sub>12</sub>			SARIMA(1,1,1)X(0,1,1) <sub>12</sub>			SARIMA(1,1,1)X(1,1,1) <sub>12</sub>		
	系数	t	P	系数	t	P	系数	t	P
AR(1)	—	—	—	0.204	1.44	0.153	0.199	1.40	0.166
SAR(12)	—	—	—	—	—	—	-0.061	-0.30	0.764
MA(1)	0.736	10.20	0.000	0.810	9.49	0.000	0.811	9.49	0.000
SMA(12)	0.556	5.59	0.000	0.590	5.95	0.000	0.547	3.02	0.003
AIC		-78.48			-78.83			-76.94	
SBC		-73.37			-71.17			-66.72	

12, SARIMA(1,1,1)X(0,1,1)<sub>12</sub>。

**2.3 模型参数估计与诊断** 比较 3 种模型的拟合效果,结果显示 ARIMA(0,1,1)X(0,1,1)<sub>12</sub> 模型最优,参数检验均有统计学意义( $P < 0.01$ ),对该模型的残差进行白噪声检验,结果在多阶延迟后, $P > 0.05$ ,说明序列信息提取充分,模型拟合良好,见表 1、2。

表 2 SARIMA(0,1,1)X(0,1,1)<sub>12</sub> 残差检验结果

延迟阶数	$\chi^2$	P
6	9.46	0.051
12	15.04	0.131
18	19.67	0.236
24	23.27	0.387

**2.4 预测检验模型效果** 根据建立的 SARIMA 模型对 2005 年 1 月至 2013 年 12 月的菌痢月发病率资料进行拟合,结果显示该模型对 2005 年 1 月至 2013 年 12 月菌痢月发病率的预测值与实际值时序曲线基本一致(图 2)。模型预测 2014 年 1~6 月的发病率数据,实际值均在 95%CI 范围之内,预测 2014 年上半年疫情呈现缓慢上升趋势,与实际情况一致。预测值与实际值的平均相对误差(MAPE)为 19%,根据预测模型理论<sup>[11-13]</sup>, $MAPE < 20\%$ ,表示预测效果良好。

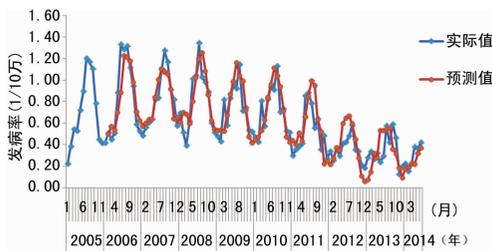


图 2 武汉市江汉区 2005 年 1 月至 2014 年 6 月菌痢预测月发病率

### 3 讨论

2005 年 1 月至 2013 年 12 月江汉区菌痢月发病率时间序列每一期的数值,都是由许多不同因素共同作用的结果,诸如人口密度、经济状况、气候及环境流行的变化往往错综复杂,因此针对单一因素的分析难以全面把握传染病的流行规律与变化趋势。然而,上述因素与时间单位(年、月、天、周)紧密关联,而时间序列模型可以在不考虑预测变量的影响因素及其关系的条件,对未来的变化进行预测,具有较高的应用价值<sup>[13-16]</sup>。其中 ARIMA 模型综合考虑趋势变化、季节和周期变动以及随机干扰的影响,是一种短期预测精确度较高的方法<sup>[7,10,13,18]</sup>。

本文利用 2005~2013 年江汉区细菌性痢疾月发病率资

料,建立 ARIMA(0,1,1)X(0,1,1)<sub>12</sub> 预测模型,各月发病率预测值的变化趋势和实际情况基本一致,效果良好,具有可行性。模型计算出的 2014 年 1 月到 6 月的发病率数据与实际值均在 95%的可信区间内,且疫情趋势走向基本相同,表示该模型对未来的情况进行了较好的跟踪与预测。疾病监测工作中,通过预测数据 95%CI 的上下限值可以判断疫情的变化规律,如果超出预测值 95%CI 的上限值,则应警惕菌痢暴发或流行的可能<sup>[7,11]</sup>。综上,通过建立合适的 ARIMA 模型预测对菌痢的风险评估具有一定的应用价值。然而,值得注意的是,一方面,影响菌痢发病的因素复杂众多,不同时间、地区的发病率存在差异,建构的模型也会存在差异,如本次研究建立模型不同于孙小宇等<sup>[16]</sup>对北京市西城区构建的 ARIMA(1,0,0)X(0,1,1)<sub>12</sub> 模型和向伦辉等<sup>[17]</sup>对上海市宝山区构建的 ARIMA(1,1,1)模型与;另一方面,ARIMA 模型只能用于短期预测,因此,随着时间的延伸,应不断加入新的实际值用以修正或拟合新的预测模型,以提高预测精度,为尽早采取针对性的防控干预措施提供科学依据<sup>[10,13,18-19]</sup>。

### 参考文献

- [1] 齐小秋,王宇,卫生部疾病预防控制局,等. 痢疾防治手册[M]. 北京:人民卫生出版社,2006:10-15.
- [2] 陈曦,洪希成,徐兴建,等. ARIMA 与 ARIMAX 模型在钉螺密度预测中的应用[J]. 公共卫生与预防医学,2014,25(2):6-9.
- [3] 杨培荣,田辉,严钊元. 自回归移动平均模型在宝鸡市细菌性痢疾月发病率预测中的应用[J]. 公共卫生与预防医学,2012,23(1):16-19.
- [4] 古万曦. 谈建立疾病预防,疫情监测控制系统[J]. 中国现代医院管理杂志,2004,2(3):71-72.
- [5] 王振家,马继胜. 新形势下公共卫生监测及其意义[J]. 地方病通报,2007,22(5):78,80.
- [6] 肖枝洪,郭明月. 时间序列分析与 SAS 应用[M]. 武汉:武汉大学出版社,2009:1-41.
- [7] 吴家兵,叶临湘,尤尔科. ARIMA 模型在传染病发病率预测中的应用[J]. 数理医药学杂志,2007,20(1):90-92.
- [8] 谭姣,雷静. ARIMA 模型在乙型病毒性肝炎发病率预测中的应用[J]. 公共卫生与预防医学,2013,24(4):8-10.
- [9] 陈超,田鑫,周剑惠,等. 流行性腮腺炎疫情时间序列模型建立与应用[J]. 中国公共卫生,2012,28(2):252-253.
- [10] 王燕. 应用时间序列分析[M]. 3 版. 北京:中国人民大学出版社,2012:1-75.
- [11] Geoge EP, Gwilym M. 时间序列分析(下转第 4128 页)

培训<sup>[12]</sup>。因此,在今后培训过程中,应灵活安排培训时间,尽可能采用短期培训的方式,在不增加应急人员工作负担的前提下,尽量提高受培训人员的比例,让更多的专业人员有更多的机会参加应急培训。

在培训师资方面,选择“国内专家”的比例相对较高,而选择高校教授及国外专家者较少,与广西卫生应急管理干部培训需求调查结果相一致,可能与国内专家具有充足的卫生应急管理理论和丰富的实践经验,而高校教授往往现场实践经验不足,国外专家不了解国内的应急管理情况有关<sup>[13]</sup>。因此,今后在培训时,在培训师资方面,应以“国内专家”为主。

在考核形式方面,71.21%的人员选择“模拟处理实际问题”,所占比例最高,体现了培训人员想通过培训来提高个人应对突发事件的能力。郭蕊等<sup>[14]</sup>认为,传统的试卷考核无法考察能力和培训效果,最好能找到多种手段结合的考核方式来综合评价培训效果。在结业方式方面,71.21%的人员认可“授予学分”的形式,女性、市级疾病预防控制机构人员更倾向于选择“授予学分”,这可能与河南省医学专业继续教育的要求及个人认同感有关。河南省继续教育要求医学专业人员每年要接受医学教育且要修够一定的学分(市级专业人员学分要求要高于县级人员),专业人员的继续教育与个人职称晋升相挂钩。男性专业人员更认可“颁发证书”的形式,认为“颁发证书”是对自己能力提升的肯定,会更有成就感。因此,应进一步健全继续教育和应急培训机制,制定完善的培训规划,出台激励措施,积极鼓励专业技术人员学习新知识、掌握新理念,提高自身素质<sup>[15]</sup>。

综上所述,如果没有开展培训需求调查,那么培训的针对性和质量肯定会受到影响。因此,在今后培训过程中,应以培训需求调查结果为依据,制订好相应的培训规划,加强培训管理,进一步健全继续教育和应急培训机制,以提高培训质量和效果,提升卫生应急处置人员的业务能力和综合素质。

## 参考文献

- [1] 王喆,梁万年,邸泽青. 疾病控制机构卫生应急培训内容与方式探讨[J]. 中国公共卫生, 2009, 25(4): 408-409.
- [2] 唐文革,杨小伶,李全乐. 培训需求分析评价及其在疾控系统的应用[J]. 现代预防医学, 2007, 34(19): 3731-3732.
- [12] Khashei M, Bijari M, Hejazi SR. Combining seasonal ARIMA models with computational intelligence techniques for time series forecasting[J]. Soft Computing, 2012, 16(6): 1091-1105.
- [13] 高蕙璇. SAS/ETS 软件使用手册[M]. 北京: 中国统计出版社, 1998: 65-106.
- [14] 杨小兵,汪鹏,江高峰. ARIMA 乘积季节模型在流行性腮腺炎发病率预测中的应用[J]. 公共卫生与预防医学, 2013, 24(6): 39-42.
- [15] 张泽武,曾耀明,卢展鹏,等. ARIMA 模型在东莞市细菌性痢疾预测中的应用[J]. 中国热带医学, 2013, 13(3):

3734.

- [3] 关莉. 卫生应急管理人员培训需求分析[J]. 中外健康文摘, 2014, 11(9): 22-23.
- [4] 中华人民共和国国家卫生和计划生育委员会. 全国卫生应急工作培训大纲(2011~2015年)[EB/OL](2011-11-21)[2015-07-29]. <http://www.moh.gov.cn/mohwsyjbg/s3581/201111/53496.shtml>.
- [5] 胡宇峰,杨荣刚,幸奠国,等. 重庆市区县卫生应急能力评估[J]. 重庆医学, 2014, 43(9): 1101-1103.
- [6] 王重建,向浩,柳东如,等. 我国突发公共卫生事件应急体系人才培养的思考[J]. 中华医学杂志, 2007, 87(35): 2455-2457.
- [7] 刘保华,吴群红,胥娇,等. 哈尔滨市基层疾控机构卫生应急能力现状分析[J]. 中国公共卫生, 2013, 29(5): 644-646.
- [8] 梁辉,马颖,秦侠,等. 卫生应急人员培训效果评估存在问题及对策[J]. 中国农村卫生事业管理, 2014, 34(4): 386-388.
- [9] 赵然,马颖,徐王权,等. 国内外卫生应急培训效果评估文献研究[J]. 中国农村卫生事业管理, 2014, 34(4): 381-385.
- [10] 关莉. 北京市传染病防控人员工作条件和培训情况分析[J]. 中外健康文摘, 2014, 11(9): 22-23.
- [11] 关丽征,王亚东,郭蕊,等. 卫生应急管理人员培训的需求分析[J]. 医学与社会, 2011, 24(2): 22-24.
- [12] 王卫军. 青海省疾病预防控制中心卫生应急工作现状调查研究[J]. 医学动物防制, 2014, 30(2): 231-233.
- [13] 周阳,刘德诚,陈发钦,等. 广西卫生应急管理干部培训需求调查[J]. 中国公共卫生, 2013, 29(11): 1641-1643.
- [14] 郭蕊,王亚东,关丽征,等. 卫生应急管理人员培训需求及模式的定性研究[J]. 医学与社会, 2011, 24(2): 4-6.
- [15] Kruk ME. Emergency preparedness and public health systems lessons for developing countries[J]. Am J Prev Med, 2008, 34(6): 529-534.

(收稿日期:2015-04-05 修回日期:2015-05-26)

(上接第 4124 页)

测与控制[M]. 顾岚,译. 北京: 中国统计出版社, 1997: 75-80.

- [12] Khashei M, Bijari M, Hejazi SR. Combining seasonal ARIMA models with computational intelligence techniques for time series forecasting[J]. Soft Computing, 2012, 16(6): 1091-1105.
- [13] 高蕙璇. SAS/ETS 软件使用手册[M]. 北京: 中国统计出版社, 1998: 65-106.
- [14] 杨小兵,汪鹏,江高峰. ARIMA 乘积季节模型在流行性腮腺炎发病率预测中的应用[J]. 公共卫生与预防医学, 2013, 24(6): 39-42.
- [15] 张泽武,曾耀明,卢展鹏,等. ARIMA 模型在东莞市细菌性痢疾预测中的应用[J]. 中国热带医学, 2013, 13(3):

327-329.

- [16] 孙小宇,初艳慧,张震,等. 3 种预测模型在北京市西城区痢疾发病数预测中的应用与比较[J]. 现代预防医学, 2014, 41(19): 3470-3474.
- [17] 向伦辉,袁国平,茅俭英. 上海市宝山区细菌性痢疾发病率的 ARIMA 模型研究[J]. 中华疾病控制杂志, 2011, 15(11): 990-992.
- [18] 张文彤. SPSS11 统计分析教程(高级篇)[M]. 北京: 北京希望电子出版社, 2002: 250-285.
- [19] 汪鹏,陈邦华,孔德广,等. 求和自回归移动平均模型在武汉市手足口病疫情预测预警中的应用[J]. 公共卫生与预防医学, 2015, 26(1): 26-29.

(收稿日期:2015-04-18 修回日期:2015-06-06)