

南通市流行性腮腺炎疫情的预测预警分析

廉丽华, 张志兰, 侯晓艳, 金红梅, 凌 剑

(南通市疾病预防控制中心, 江苏 南通 226007)

[摘要] 目的:探讨时间序列分析中的求和自回归移动平均模型(autoregressive integrated moving average, ARIMA)乘积季节模型在南通市流行性腮腺炎疫情预测预警中的应用。方法:收集“中国疾病预防控制中心”中 2004 年 1 月~2013 年 11 月流行性腮腺炎的月发病监测数据,建立时间序列数据库,对每月腮腺炎的发病人数进行 ARIMA 模型拟合,利用模型对 2004 年 1 月~2013 年 11 月的数据进行回代预测,并对 2013 年 12 月~2014 年 6 月各月的流行性腮腺炎的发病情况进行前瞻性预测。结果:构建流行性腮腺炎的 ARIMA(0,1,2)(1,1,0)₁₂ 模型为 $(1-B)(1+0.51B^{12})case_t = (1-0.24B-0.29B^2)(1-B^{12})\varepsilon_t$, 其中 B 代表后移算子, $case_t$ 代表年腮腺炎发病数, ε_t 为随机误差。回代预测的实际值和预测值基本相符,均在 95%可信区间内。前瞻性预测结果符合流行性腮腺炎的流行特征。结论:该模型能较好模拟并预测腮腺炎的发病情况,流行性腮腺炎发病数有增加趋势,应进一步分析本地腮腺炎的流行特征及其免疫策略。

[关键词] 时间序列分析; ARIMA 模型; 腮腺炎; 预测预警

[中图分类号] R181.8

[文献标志码] B

[文章编号] 1007-4368(2014)11-1618-03

doi: 10.7655/NYDXBNS20141138

流行性腮腺炎是由流腮病毒引起的急性呼吸道传染病,2004 年作为丙类传染病纳入法定传染病报告并实施网络直报。随着免疫规划实施相关疾病得到有效控制,腮腺炎的发病已居南通市疫苗可预防传染病之首。由于腮腺炎好发于儿童和青少年,本市的托幼机构和学校每年都会发生腮腺炎的流行和爆发,而目前尚未开展针对腮腺炎的监测工作。因此,对腮腺炎的发病情况及时进行预测,有助于控制腮腺炎的流行和爆发。本文通过分析南通市 2004~2013 年的流行性腮腺炎的月发病资料,运用时间序列分析中的求和自回归移动平均模型(Autoregressive Integrated Moving Average, ARIMA)乘积季节模型对 119 个月的病例数据进行建模拟合,并预测了 2013 年 12 月~2014 年 6 月的发病情况。

1 资料和方法

1.1 资料

本研究中南通市历年腮腺炎发病数资料(2004 年 1 月~2013 年 11 月)来源于中国疾病预防控制中心信息系统 V2.0。

1.2 方法

1.2.1 模型简介

ARIMA 乘积季节模型 $(p, d, q) \times (P, D, Q)_s$ 记为: $\Phi(B)U(B^s)\nabla^d\nabla^p X_t = V(B^s)\Theta(B)\varepsilon_t$, 其中, t 代表时间, X_t 表示相应时间序列, ∇ 表示差分算子, $\nabla = X_t - X_{t-1}$,

∇^d 表示经过连续 d 次差分, B 是后移算子, $\nabla = 1 - B$, p, d, q 分别表示自回归阶数、差分阶数和移动平均阶数; $\Phi(B)$ 表示自回归算子, $\Theta(B)$ 表示移动平均算子; ε_t 代表独立扰动或者随机误差^[1], s 表示一个季节循环中观测的个数, $\Phi(B)\nabla^d X_t$ 表示同一周期内不同周期点的相关关系, $U(B_s)\nabla^p$ 表示不同周期的同一周期点上的相关关系。当 $P=D=Q=0$ 时该模型就是一般的 ARIMA 模型^[2]。

1.2.2 建立模型

以南通市 2004 年 1 月~2013 年 11 月份流行性腮腺炎发病数据为基础建立数据库,应用 SAS9.1.3 统计软件进行数据处理与分析,通过如下步骤建立 ARIMA 乘积季节模型^[2]。数据预处理:通过自相关系数或者平方根检验法判断其平稳性,若该数列的平均值和方差始终为常数,则称其为平稳,否则需要进行差分或者对数转换等方法使其平稳化。模型的识别、定阶和参数估计:ARIMA 乘积季节模型的阶数识别和定阶采用了 Box-Jenkins 法,参数估计采用最小二乘或非线性估计法。模型拟合优度检验:对观测值和模型拟合值的残差进行分析,若残差序列是白噪声序列,则所建模型为最终模型,否则需要重新建立模型,重复上述步骤,直到残差序列为白噪声为止。

1.2.3 预测

利用所建立的 ARIMA 乘积季节模型对南通市 2004 年 1 月~2013 年 11 月月发病数进行回代预

测,并预测 2013 年 12 月~2014 年 6 月月发病情况。

2 结果

2.1 流行性腮腺炎的发病趋势分析

由图 1 可见,南通市 2004 年 1 月~2013 年 11 月流行性腮腺炎的月发病序列呈现明显的非平稳性、周期性和季节性。需对其进行差分以平稳时间序列。

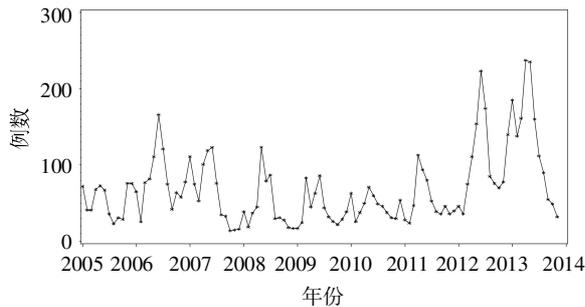


图 1 南通市 2004~2013 年 11 月流行性腮腺炎月发病原始时间序列图

2.2 建立模型

2.2.1 数据平稳化

对流行性腮腺炎月发病序列进行季节差分和一阶差分后,如图 2 所示,经过 ADF 平方根检验 $P < 0.000 1$,序列已达到平稳。

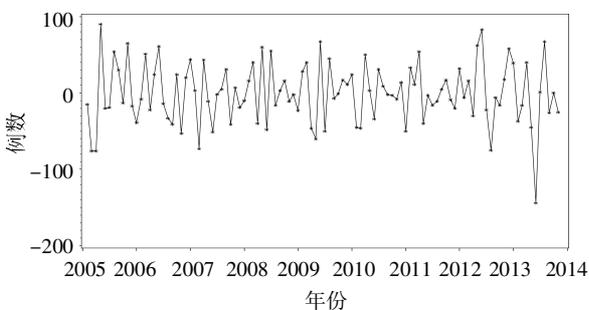


图 2 一阶差分和季节差分后时间序列图

2.2.2 模型识别、定阶和参数估计

表 1 乘积季节 ARIMA 模型参数估计结果

参数	滞次	估计值	标准误(S_{ξ})	t 值	P 值
MA(1,1)	1	0.238 79	0.095 46	2.50	0.013 9
MA(1,2)	2	0.288 16	0.097 78	2.95	0.004 0
AR(1,1)	12	-0.511 81	0.101 56	-5.04	<0.000 1

然而,腮腺炎的流行是多种因素共同影响的结果,且各因素之间的关系错综复杂,很难运用一般的因果模型来预测。在总人口基数没有太大波动时,利用腮腺炎发病人数序列自身的变动规律,建立相应的动态模型可有效对腮腺炎疫情进行预测预警^[4-5]。

将各种复杂因素的综合效应统一蕴涵于时间变量中,这是时间序列分析应用于疾病预测的一个突

通过观察 1 阶差分和季节差分后的平稳序列的自相关图和偏自相关图考虑使用 ARIMA(0,1,2)(1,1,0)₁₂ 模型拟合原始序列。应用 SAS 9.1.3 统计软件进行参数估计,结果见表 1。根据参数估计的结果,乘积季节 ARIMA 模型可表示为: $(1-B)(1+0.51B^{12})case_t = (1-0.24B-0.29B^2)(1-B^{12})\varepsilon_t$, 其中 B 代表后移算子, X_t 代表年腮腺炎发病数, ε_t 为随机误差。

2.2.3 模型的拟合优度检验

对残差序列进行检验,结果如表 2 所示。拟合检验统计量的 P 值都大于检验水准 0.05, 可以认为该残差序列已为白噪声序列,说明信息已被充分提取,即所建模型为最终模型。

2.3 模型预测

应用所建模型对 2004 年 1 月~2013 年 11 月各月腮腺炎发病数进行回代预测,并预测 2013 年 12 月~2014 年 6 月各月的发病情况。如图 3 所示,2004 年 1 月~2013 年 11 月的观察值和预测值基本相符,95%可信区间也较窄,说明该模型能较好模拟该时期腮腺炎的发病情况。对 2013 年 12 月~2014 年 6 月的发病情况预测显示,2014 年的腮腺炎发病趋势将同 2013 年一样,呈现季节性变化,发病总数较 2004~2011 年有上浮趋势(图 3,表 3)

3 讨论

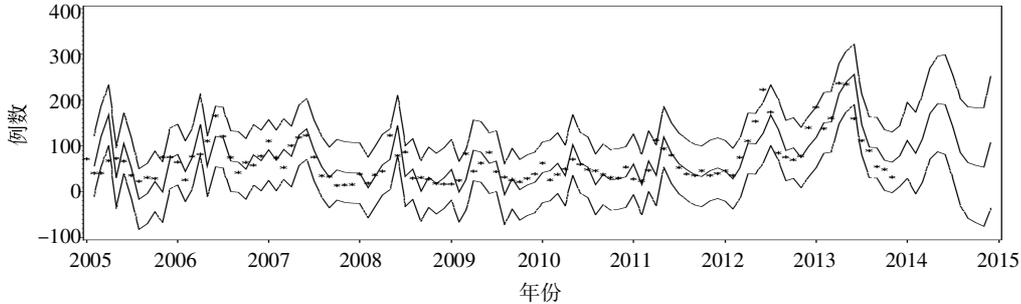
流行性腮腺炎是一种自限性疾病,但流腮病毒不仅仅侵犯腮腺,还会累及其他许多腺体和器官。山东省对 2001~2003 年腮腺炎临床诊断病例的调查分析显示,33.30%的调查病例有脑膜炎、脑炎、胰腺炎、睾丸炎、心肌炎和中耳炎等并发症^[3]。由于腮腺炎尚未纳入疾病监测,及时准确地对腮腺炎疫情进行预测,在当前腮腺炎的预防控制上显得尤为重要。

出优点^[6-7]。传统的时间序列模型要求序列具有平稳的线性趋势,但实际上腮腺炎的发病情况有着明显的季节性和周期性,如果不考虑这些因素的影响,做出的预测往往不准确。本研究采用 ARIMA 乘积季节模型考虑了季节周期因素、趋势因素和随机误差等对平稳性造成的影响,因而改善了预测效果^[8-10]。

对南通市 2004~2013 年流行性腮腺炎的月发

表2 乘积季节ARIMA模型残差序列白噪声检验结果

延迟阶数	χ^2 值	P 值	残差自回归系数
6	5.93	0.115 3	0.023, 0.041, 0.006, -0.014, -0.223, -0.019
12	10.09	0.343 1	-0.030, 0.010, -0.049, 0.146, -0.011, -0.099
18	14.84	0.463 1	0.054, -0.088, -0.016, -0.162, -0.016, -0.008
24	20.03	0.519 0	0.004, 0.009, 0.078, -0.091, 0.042, -0.147



图中星号代表实际值,中间曲线代表预测值,上下曲线为95%可信区间。
图3 2004~2014年南通市流行性腮腺炎月发病模型预测图

表3 2013年12月~2014年12月模型预测值及其95%可信区间 (例)

时间	预测值(95%CI)
2013年12月	77.83 (11.60~144.05)
2014年1月	112.19 (28.96~195.42)
2014年2月	84.13 (-4.80~173.05)
2014年3月	115.31 (21.03~209.60)
2014年4月	170.84 (71.49~270.20)
2014年5月	191.87 (87.70~296.05)
2014年6月	190.57 (81.79~299.36)

病数分析发现,南通市腮腺炎发病呈现明显的季节性,冬春季节高发,2004~2011年较平稳,2012年开始数据显示上升趋势,这与实际情况相符。由于腮腺炎发病的影响因素很多,如流动人口、卫生条件、疫苗接种率、人群易感性等,都会对预测效果产生影响。本研究在ARIMA乘积季节模型在腮腺炎发病预测的可行性基础上,对腮腺炎的发病趋势进行短期预测,结果提示腮腺炎有逐年增加的趋势。下一步我们将对腮腺炎的流行特征和现接种疫苗株以及免疫接种方案等可控因素进行进一步调查研究^[10],采用可行防控措施将疫情控制在更低水平。

[参考文献]

[1] Yong MH, Yoon-Jin Lee. Detecting misspecifications in autoregressive conditional duration models and non-negative time series processes [J]. J Time Ser Anal, 2011, 32(1):

1-32

[2] 王燕. 应用时间序列分析[M]. 2版. 北京:中国人民大学出版社, 2008: 165-170

[3] 许青, 徐爱强, 宋立志, 等. 山东省2001~2003年流行性腮腺炎临床诊断病例流行病学调查分析[J]. 中国计划免疫, 2006, 12(6): 497-499

[4] 朱猛, 祖荣强, 霍翔, 等. 时间序列分析在流感疫情预测预警中的应用[J]. 中华预防医学杂志, 2011, 45(12): 1108-1111

[5] 王超, 丁勇, 陆群, 等. ARIMA乘积季节模型在我国甲肝发病预测中的应用[J]. 南京医科大学学报:自然科学版, 2014, 34(1): 75-79

[6] 丁晓艳, 彭志行, 陶红, 等. 运用时间序列模型对麻疹流行趋势的预测和分析[J]. 南京医科大学学报:自然科学版, 2011, 31(8): 1200-1202

[7] 彭志行, 陶红, 贾成梅, 等. 时间序列分析在麻疹疫情预测预警中的应用研究[J]. 中国卫生统计, 2010, 27(5): 459-463

[8] 邓丹, 王润华. 时间序列分析及其在卫生事业中的应用[J]. 数理医药学杂志, 2002, 15(5): 455-457

[9] 彭志行, 鲍昌俊, 赵杨, 等. ARIMA乘积季节模型及其在传染病发病预测中的应用[J]. 数理统计与管理, 2008, 27(2): 362-368

[10] 张冬艳, 冯燕, 钟淑玲, 等. 2005—2010年浙江省腮腺炎流行株与疫苗株S79全基因组序列比较分析[J]. 中华预防医学杂志, 2012, 46(3): 252-257

[收稿日期] 2014-02-11