

ARIMA 乘积季节模型在陕西省手足口病预测中的应用

刘峰 朱妮 邱琳 王敬军 王维华

710054 西安,陕西省疾病预防控制中心(刘峰、王敬军),信息所(朱妮、邱琳),传染病预防控制所(王维华)

通信作者:王维华, Email: myjshy@163.com

DOI:10.3760/cma.j.issn.0254-6450.2016.08.013

【摘要】 目的 应用自回归求和移动平均 (ARIMA) 乘积季节模型预测陕西省手足口病发病人数。方法 利用 R 软件对 2009 年 1 月至 2015 年 6 月手足口病数据构建 ARIMA 乘积季节模型,同时利用 2015 年 7—12 月实际发病人数与模型拟合数据比较,评价模型的预测性能,预测 2016—2017 年的手足口病发病人数。结果 陕西省手足口病发病人数具有明显的季节性,建立了 ARIMA (2, 1, 0) × (1, 1, 0)₁₂ 乘积季节模型,预测方程为 $(1 - B)(1 - B_{12}) \ln(X_t) = \frac{(1 - 1.000B)}{(1 - 0.532B - 0.363B^2)(1 - 0.644B_{12} - 0.454B_{12}^2)} \varepsilon_t$, 与 2015 年 7—12 月实际发病人数比较,绝对误差的平均值为 531.535,相对误差的平均值为 0.114,预测 2016—2017 年陕西省手足口病发病人数与 2015 年基本保持一致。结论 ARIMA(2,1,0) × (1,1,0)₁₂ 乘积季节模型可用于陕西省手足口病的发病人数预测。

【关键词】 手足口病; 自回归求和移动平均乘积季节模型; 预测

Application of R-based multiple seasonal ARIMA model, in predicting the incidence of hand, foot and mouth disease in Shaanxi province Liu Feng, Zhu Ni, Qiu Lin, Wang Jingjun, Wang Weihua

Shaanxi Provincial Center for Disease Control and Prevention, Xi'an 710054, China (Liu F, Wang JJ); Institute for Health Information of Shaanxi Provincial Center for Disease Control and Prevention, Xi'an 710054, China (Zhu N, Qiu L); Institute for Communicable Disease Control and Prevention of Shaanxi Provincial Center for Disease Control and Prevention, Xi'an 710054, China (Wang WH)

Corresponding author: Wang Weihua, Email: myjshy@163.com

【Abstract】 Objective To apply the 'auto-regressive integrated moving average product seasonal model' in predicting the number of hand, foot and mouth disease in Shaanxi province.

Methods In Shaanxi province, the trend of hand, foot and mouth disease was analyzed and tested, under the use of R software, between January 2009 and June 2015. Multiple seasonal ARIMA model was then fitted under time series to predict the number of hand, foot and mouth disease in 2016 and 2017. **Results** Seasonal effect was seen in hand, foot and mouth disease in Shaanxi province. A multiple seasonal ARIMA (2,1,0) × (1,1,0)₁₂ was established, with the equation as $(1 - B)(1 - B_{12}) \ln(X_t) = \frac{(1 - 1.000B)}{(1 - 0.532B - 0.363B^2)(1 - 0.644B_{12} - 0.454B_{12}^2)} \varepsilon_t$.

The mean of absolute error and the relative error were 531.535 and 0.114, respectively when compared to the simulated number of patients from Jun to Dec in 2015. Results under the prediction of multiple seasonal ARIMA model showed that the numbers of patients in both 2016 and 2017 were similar to that of 2015 in Shaanxi province. **Conclusion** Multiple seasonal ARIMA (2,1,0) × (1,1,0)₁₂ model could be used to successfully predict the incidence of hand, foot and mouth disease in Shaanxi province.

【Key words】 Hand foot and mouth disease; Multiple seasonal ARIMA model; Prediction

近年来,手足口病的暴发或流行,引起儿童死亡,受到越来越多的关注^[1-2]。2008年5月将手足口病纳入丙类法定传染病管理以来,陕西省报告的发病数和死亡数逐年上升。针对陕西省手足口病的研

究主要集中在病原学及流行特征分析^[3-4],本研究以 R 软件为基础,应用自回归求和移动平均 (autoregressive integrated moving average, ARIMA) 乘积季节模型,对陕西省手足口病发病人数进行分

析预测,以了解陕西省手足口病发病人数的变化规律和发展趋势。

资料与方法

1. 数据来源:2009—2015 年陕西省各级各类医疗卫生机构通过《传染病报告管理信息系统》报告的手足口病发病人数。2009 年 1 月至 2015 年 6 月手足口病发病人数数据用于模型拟合,2015 年 7 月至 2015 年 12 月手足口病发病人数数据用于模型预测效果评价。

2. 研究方法:ARIMA 模型是一种基于时间序列分析和预测的方法。随机季节模型与 ARIMA 模型结合即为 ARIMA 季节模型,一般表示为 $ARIMA(p, d, q) \times (P, D, Q)_s$, 各参数的意义:非季节自回归阶数 p 、非季节差分阶数 d 、非季节移动平均阶数 q 、季节自回归阶数 P 、季节差分阶数 D 和季节移动平均阶数 Q , 季节长度 s 。当 $P=D=Q=0$ 时,该模型为 ARIMA 模型。ARIMA 乘积季节模型能综合考虑季节、趋势和随机干扰等因素,对时间序列预测效果较好^[5]。此次研究采用 R-3.2.2 软件进行建模及数据处理分析。建模过程:①序列平稳:通过“tseries”软件包中的“ts”函数将 2009 年 1 月至 2015 年 6 月手足口病发病人数定义为以月为单位的时间序列,利用“decompose”函数绘制并分解时间序列,观察时间序列的趋势及周期变化,然后通过“diff”函数对序列进行差分,使之成为零均值、无明显趋势变化的平稳序列。②模型识别:对平稳时间序列,分别通过“acf”和“pacf”函数观察序列的自相关系数 (autocorrelation function, ACF) 和偏自相关系数 (partial autocorrelation function, PACF) 初步确定 p, d, q 和 P, D, Q 的阶值。若自相关函数为拖尾,且偏相关函数为截尾,适合 AR 模型;若自相关函数为截尾,且偏相关函数为拖尾,适合 MA 模型;若自相关函数和偏相关函数均为拖尾,则序列适合 ARIMA 模型。参数 P, Q 超过 2 阶的情况不多见,可分别取 0、1、2 并由低阶到高阶逐个试验,根据模型的拟合优度等指标进行判断,以确定这两个参数。 s 根据疾病的背景知识获得。③参数估计:ARIMA 乘积季节模型运用最大似然估计 (maximum likelihood estimation, MLE) 或最小二乘估计 (least squares estimation, LSE) 法,根据 p, d, q 和 P, D, Q 的阶值计算出自回归和移动平均系数。④模型诊断:通过“Box.test”函数计算 Box-Ljung 统计量,如果统计量差异无统计学意义,可认为残差为白噪声,若残差序

列不是白噪声,则需对模型进行改进。若残差序列满足白噪声过程,通过 AIC 等指标挑选最优模型, AIC 越小,模型的拟合程度越好。此外,通过比较 2015 年 7—12 月手足口病实际发病人数和模型拟合发病人数,计算各月绝对误差和相对误差水平来评价模型的拟合效果。⑤模型预测:以上步骤建立 2009 年 1 月至 2015 年 6 月陕西省手足口病发病人数模型,通过“forecast.Arima”函数模型预测 2016—2017 年发病人数。

3. 统计学分析:本研究采用 Excel 软件建立数据库,用 R-3.2.2 中“tseries”和“forecast”软件包进行数据处理和建立模型。R 软件是由奥克兰大学的 Robert Gentleman 和 Ross Ihaka 及其他志愿人员开发的统计分析软件,可用来探索数据、统计分析及绘图等。

结 果

1. 发病人数变化趋势:2009 年 1 月至 2015 年 6 月陕西省手足口病年平均发病人数为 52 631.14 例,月平均发病人数 4 385.93 例,其中发病人数最多的是 2015 年 5 月,共发病 18 074 例。由时间序列可看出陕西省 7 年间手足口病发病人数存在一定的周期性,将时间序列用“decompose”语句按总体趋势、季节趋势及随机误差分解后可看出,总体趋势除在 2011 年出现下降外,2009—2015 年手足口病发病人数呈整体上升趋势,提示该序列为非平稳时间序列。此外,手足口病存在明显的季节性,春夏之交出现高峰,两头下降,随机误差保持在一定水平(图 1)。

2. ARIMA 乘积季节模型构建:

(1)序列平稳化:由图 1 可知时间序列呈现明显的季节变化。用“diff”语句进行一阶 12 步季节差分消除时间序列趋势和季节影响,得到了一个近似平稳的随机序列(图 2),符合 ARIMA 乘积季节模型对平稳性的要求。

(2)模型识别:根据差分变换的次数,初步确定以 12 个月为周期的 $ARIMA(p, 1, q)(P, 1, Q)_{12}$ 乘积季节模型,分别用“acf”和“pacf”语句画出自相关图(图 3)和偏自相关图(图 4)。由图可知,自相关系数在 0 阶截尾,12 阶、24 阶有明显波动,说明差分后序列中含有显著地季节效应,偏自相关系数在 2 阶截尾,同时在滞后 12 阶显著大于可信区间范围,也显示出明显的季节周期,考虑模型的 AR 因子为 $(1 - \theta_1 B - \theta_2 B^2)\varepsilon_t$ 。根据上述特征判断 $p=2, q=0$, 但 P 和 Q 的判断比较困难,根据《现代医学统计学》(方积乾, 陆

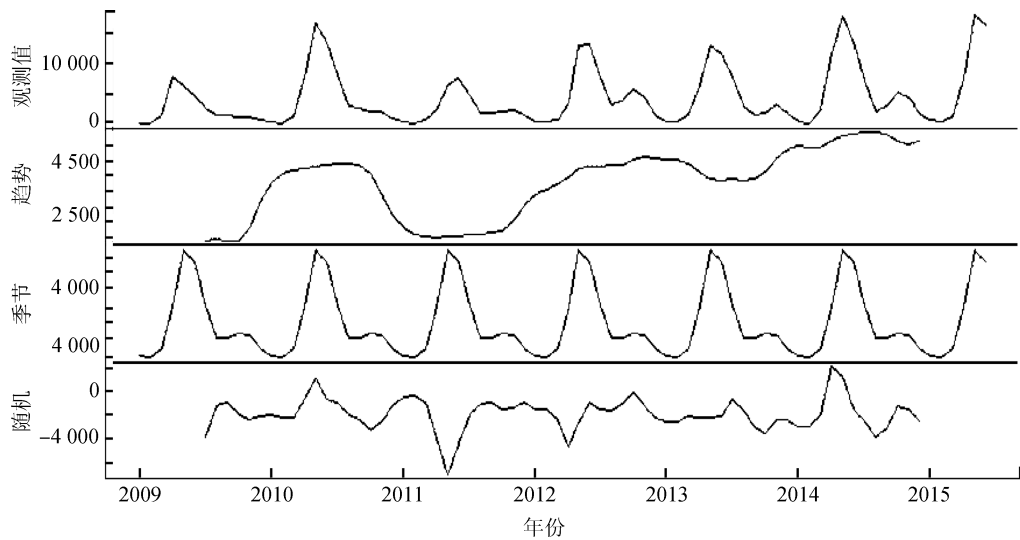


图1 2009年1月至2015年6月陕西省手足口病发病人数时间序列

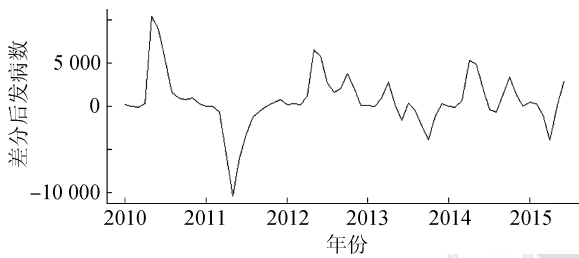


图2 经过一阶12步季节差分后的手足口病发病数序列图

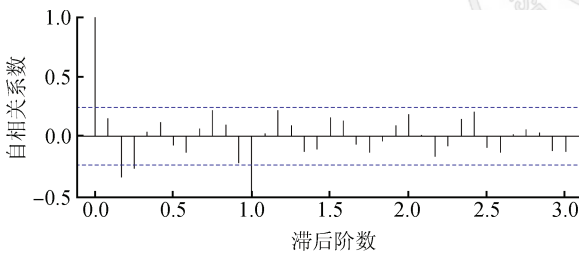


图3 变换后序列自相关图

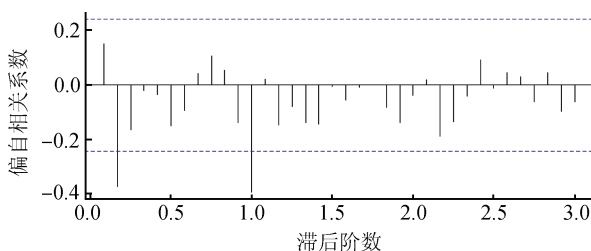


图4 变换后序列偏自相关图

盈. 人民卫生出版社, 2002), P和Q超过2阶的情况很少见, 可分别取0、1、2逐个试验, 根据参数及模型总体的显著性、拟合优度等指标进行比较, 选择最佳模型。

(3)参数估计和模型诊断: 拟合的3个备选模型参数估计值及检验结果见表1。残差白噪声检验和拟合优度检验用来评估各个备选模型对时间序列内

在综合信息的提取程度及各个备选模型之间的优劣比较。用“Box.test”语句“Ljung-Box”统计量对3个备选模型的残差值进行检验, 残差均为白噪声, 由AIC统计量可知, 模型1拟合效果最好(表1)。综合以上结论, 选择模型ARIMA(2, 1, 0) × (1, 1, 0)₁₂是合适的。

表1 备选模型的参数估计及检验

参数	ARIMA(2, 1, 0) × (1, 1, 0) ₁₂		ARIMA(2, 1, 0) × (1, 1, 1) ₁₂		ARIMA(2, 1, 0) × (2, 1, 1) ₁₂	
	估计值	s _r	估计值	s _r	估计值	s _r
AR1	0.142	0.122	0.140	0.122	0.139	0.123
AR2	-0.346	0.119	-0.333	0.124	-0.333	0.124
SAR1	-0.622	0.114	-0.703	0.197	-0.893	0.856
SAR2	-	-	-	-	-0.124	0.567
SMA1	-	-	0.139	0.310	0.321	0.828
AIC	1 067.14	-	1 068.99	-	1 070.93	-
残差序列L-B检验的P值	0.667	-	0.673	-	0.671	-

$$(1-B)(1-B_{12})\text{Ln}(X_t) = \frac{1}{(1 + 0.142B - 0.346B^2)(1 - 0.622B_{12})} \varepsilon_t$$

3. 模型评价: 利用ARIMA(2, 1, 0) × (1, 1, 0)₁₂乘积季节模型对陕西省2015年7—12月手足口病发病人数进行预测(表2)。该模型预测值的绝对误差、相对误差范围分别为146.520~1 843.137和

表2 2015年7—12月手足口病逐月发病人数的预测结果

月份	预测值	实际值	绝对误差	相对误差
7	7 968.863	9 812	1 843.137	0.188
8	2 930.520	2 784	146.520	0.053
9	2 382.520	2 026	356.520	0.176
10	3 433.454	3 260	173.453	0.053
11	4 065.724	4 349	283.276	0.065
12	2 230.700	2 617	386.300	0.148

0.053~0.188, 绝对误差的平均 531.535, 相对误差平均 0.114, 提示 ARIMA(2, 0, 0) × (1, 1, 0)₁₂ 模型具有较佳的预测性能。

4. 模型预测: 用“forecast”函数包中的“plot.forecast”过程, 绘制模型拟合的陕西省 2009年1月至2015年12月手足口病发病人数(虚线)及 2016—2017年手足口病发病人数(实线)序列图(图5)。提示 2016年和2017年陕西省手足口病发病水平与2015年接近。

讨 论

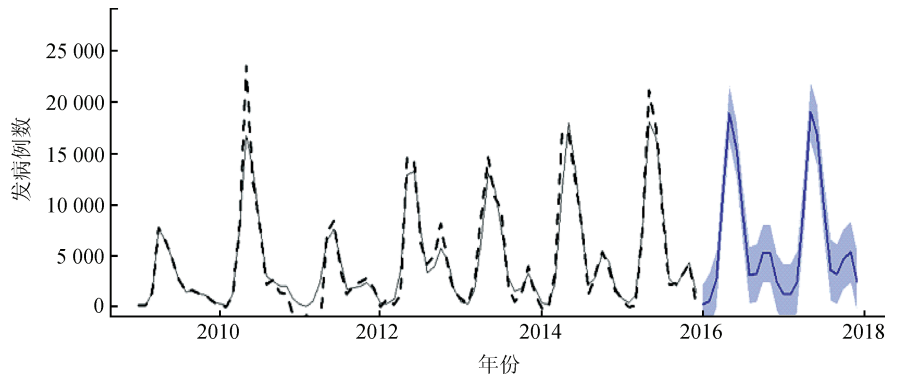
自 2009 年开展监测以来, 手足口病发病人数不断呈上升趋势, 而发病人数的预测对该病的防治工作具有重要指导意义。数学模型预测是进行手足口病预警及控制的较好手段。手足口病预测模型主要建立在该病疫情发展的基础上。目前手足口病的预测已有多种数学模式^[6-8], 但均需要传播环节或敏感人群的分布参数。一旦参数发生变化或无法获得时, 也将无法做出准确预测。此外, 实际工作中, 仍有很多影响手足口病的因素未被考虑到模型中, 从而影响到结果的准确性和可延性。而时间序列预测不需考虑实际参数, 而是在长期的实际情况下达到平衡状态时做出预测^[9-10], 手足口病的发病人数序列具有非线性和季节性, 给 P, Q 的取值带来困难。

本研究用开源的 R 软件以及相关软件包^[11], 通过 ARIMA 乘积季节模型分析 2009 年 1 月至 2015 年 6 月陕西省手足口病的时间序列数据, 探讨该病的发病规律, 并对发病人数作出预测。该模型可为陕西省手足口病预防措施的制定提供一定依据。作为一种数据处理方法, 时间序列预测方法与其他常用的预测方法一样, 主要从数理上反映疾病的统计规律, 作为疾病防治的决策还须考虑其他因素对预测结果的影响。另外, 本文提出的 ARIMA 乘积季节模型能预测不同时间的手足口病发病人数, 但对短期预测还是长期预测较为准确, 还需进一步研究和后续数据的支持。

利益冲突 无

参 考 文 献

[1] 孙军玲, 张静. 手足口病流行病学研究进展[J]. 中华流行病学杂志, 2009, 30(9): 973-976. DOI: 10.3760/cma.j.issn.0254-6450.2009.09.027.



注: 虚线部分为模型拟合值, 实线部分为模型预测值

图5 2016—2017年陕西省手足口病发病人数预测

- Sun JL, Zhang J. A review on the advancement of epidemiology on hand-foot-mouth disease[J]. Chin J Epidemiol, 2009, 30(9): 973-976. DOI: 10.3760/cma.j.issn.0254-6450.2009.09.027.
- [2] 杨智宏, 朱启镛, 李秀珠, 等. 2002年上海儿童手足口病病例中肠道病毒 71 型和柯萨奇病毒 A 组 16 型的调查[J]. 中华儿科杂志, 2005, 43(9): 648-652. DOI: 10.3760/cma.j.issn.0578-1310.2005.09.003.
- Yang ZH, Zhu QR, Li XZ, et al. Detection of enterovirus 71 and coxsackievirus A16 from children with hand, foot and mouth disease in Shanghai, 2002 [J]. Chin J Pediatr, 2005, 43(9): 648-652. DOI: 10.3760/cma.j.issn.0578-1310.2005.09.003.
- [3] 郭晓荣, 郁会莲, 邱琳, 等. 2010年陕西省手足口病报告发病高峰期重症及死亡病例情况分析[J]. 疾病监测, 2011, 26(8): 608-610. DOI: 10.3784/j.issn.1003-9961.2011.08.007.
- Guo XR, Yu HL, Qiu L, et al. Severe and fatal cases of hand foot and mouth disease during incidence peak period in Shaanxi province, 2010 [J]. Dis Surveill, 2011, 26(8): 608-610. DOI: 10.3784/j.issn.1003-9961.2011.08.007.
- [4] 郁会莲, 郭晓荣, 邱琳, 等. 2008—2011年陕西省手足口病流行特征分析[J]. 疾病监测, 2012, 27(6): 443-445. DOI: 10.3784/j.issn.1003-9961.2012.6.008.
- Yu HL, Guo XR, Qiu L, et al. Epidemiology of hand foot and mouth disease in Shaanxi, 2008-2011 [J]. Dis Surveill, 2012, 27(6): 443-445. DOI: 10.3784/j.issn.1003-9961.2012.6.008.
- [5] 于林凤, 吴静, 周锁兰, 等. ARIMA 季节模型在我国丙肝发病预测中的应用[J]. 郑州大学学报: 医学版, 2014, 49(3): 344-348. DOI: 10.13705/j.issn.1671-6825.2014.03.014.
- Yu LF, Wu J, Zhou SL, et al. Application of seasonal ARIMA model in forecasting incidence of hepatitis C in China [J]. J Zhengzhou Univ: Med Sci, 2014, 49(3): 344-348. DOI: 10.13705/j.issn.1671-6825.2014.03.014.
- [6] 王梦圆, 马文文, 孙牧, 等. 手足口病传播的数学模型研究[J]. 生物医学工程研究, 2014, 33(1): 35-38.
- Wang MY, Ma WW, Sun M, et al. Research for transmission of hand-foot-mouth disease based on mathematical model [J]. J Biomed Eng Res, 2014, 33(1): 35-38.
- [7] 李春. 手足口病传播的数学建模与研究[D]. 信阳: 信阳师范学院, 2011. DOI: 10.7666/d.y1946660.
- Li C. The modeling and analysis for the transmission of hand foot mouth disease [D]. Xinyang: Xinyang Normal University, 2011. DOI: 10.7666/d.y1946660.
- [8] 黄英芬. 手足口病数学模型的研究现状[J]. 科技信息, 2013, (20): 139. DOI: 10.3969/j.issn.1001-9960.2013.20.119.
- [9] 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: 135-152.
- Wang Y. Applied Time Series Analysis [M]. 3rd edition. Beijing: China Renmin University Press, 2012: 135-152.
- [10] 刘刚, 唐宋, 孙文杰. 时间序列分析法在香港结核病预测中的应用[J]. 中国卫生统计, 2012, 29(2): 226-228. DOI: 10.3969/j.issn.1002-3674.2012.02.021.
- Liu G, Tang S, Sun WJ. Application of time series analysis in the prediction of tuberculosis in Hong Kong [J]. Chin J Health Stat, 2012, 29(2): 226-228. DOI: 10.3969/j.issn. 1002-3674.2012.02.021.
- [11] Forecasting using ARIMA or ARFIMA models [DB/CD] (2016-03-30). <http://finzi.psych.upenn.edu/R/library/forecast/html/forecast.Arima.html>.

(收稿日期: 2016-03-11)

(本文编辑: 王岚)