临床研究

DOI:10.13406/j.cnki.cyxb.001372

ARIMA乘积季节模型在重庆市流行性乙型脑炎预测中的应用

周春碚，姚　宁
（重庆市疾病预防控制中心，重庆　400042）

【摘　要】**目的**：应用自回归求和移动平均（autoregressive integrated moving average，ARIMA）乘积季节模型预测重庆市流行性乙型脑炎（简称乙脑）发病数。**方法**：利用 R 软件对重庆市 2006 年 1 月到 2015 年 6 月乙脑报告病例数进行 ARIMA 模型建模拟合，选择预测模型进行相互比较确定最优模型。用 2015 年 7 至 12 月实际报告病例数与拟合值的比较来评价模型的预测效果，并对 2016 至 2017 年重庆市乙脑报告发病数进行预测。**结果**：重庆市乙脑发病人数呈逐年下降趋势，报告病例具有明显季节分布特征，ARIMA$(0,0,1)\times(1,1,2)_{12}$ 模型很好地拟合了时间序列，该模型赤池信息量准则（Akaike information criterion，AIC）、许瓦兹贝叶斯准则（Schwarz Bayesian criterion，SBC）值均最小且预测值与实际值的平均相对误差为 0.12，平均绝对百分比误差为 7.81%。进一步用该模型预测重庆市 2016 至 2017 年乙脑病例数分别为 35 例和 32 例，发病高峰仍是 7 至 8 月。**结论**：利用ARIMA乘积季节模型对乙脑发病数拟合较好，短期预测结果良好；与 2015 年比较，预测 2016 至 2017 年乙脑报告发病数略微减少。

【关键词】流行性乙型脑炎；自回归求和移动平均乘积季节模型；预测
【中图分类号】R512.32　　【文献标志码】A　　【收稿日期】2016-12-20

Application of autoregressive integrated moving average model(ARIMA) in the prediction of Japanese encephalitis in Chongqing

Zhou Chunbei, Yao Ning
(*Chongqing Center for Disease Control and Prevention*)

【Abstract】***Objective***: To predict the possible Japanese encephalitis (JE) cases in Chongqing using autoregressive integrated moving average model (ARIMA). ***Methods***: An ARIMA prediction model was established by R software using data of reported JE cases in Chongqing from January 2006 to June 2015. Comparison between predicted numbers and reported cases from July to December of 2015 was examined for model assessment, then the most fitted model was used to predict JE cases in year 2016-2017. ***Results***: Reported JE cases in Chongqing decreased annually, and fit the characteristic of seasonal change. ARIMA$(0,0,1)\times(1,1,2)_{12}$ can appropriately fit the time series with the smallest value of Akaike information criterion (AIC) and Schwarz Bayesian criterion (SBC) index with a mean relative residual of 0.12, the mean absolute percentage error (MAPE) was 7.81%. The predicted JE cases in year 2016 and 2017 were 35 and 32 respectively, the incidence peak was still at July and August. ***Conclusion***: The ARIMA model can be used to fit the time series trend of JE disease, predicted JE cases in year 2016 and 2017 would be less than that of reported cases in 2015.

【Key words】Japanese encephalitis; multiple seasonal ARIMA model; prediction

流行性乙型脑炎是由乙脑病毒引起的急性虫媒传染病[1]，主要以损害中枢神经系统为主，感染者大部分是 14 岁以下儿童[2]且预后较差[3]。重庆市是乙脑的流行地区，在纳入扩大免疫规划（expanded program on immunization，EPI）后报告病例数逐年降低。乙脑在夏季高发，与蚊媒密度、气象因素密切相关，因此对乙脑的预测难度较大[4]。ARIMA 模型法是最常用的时间序列预测法，并已广泛应用于疾病预测领域且取得了不错的预测效果[5-9]。本文应用 R 软件中的自回归求和移动平均乘积季节模型，对 2006 至 2015 年重庆市乙脑报告发病数进行拟合分析和预测，了解重庆市乙脑发病人数的变化规律和今后的发病趋势。

作者介绍：周春碚，Email：522576906@qq.com，研究方向：疾病控制。
通信作者：姚　宁，Email：605329038@qq.com。
基金项目：重庆市卫生计生委科研面上资助项目（编号：2015MSXM091）。
优先出版：http://kns.cnki.net/kcms/detail/50.1046.R.20170630.1042.034.html（2017-06-30）

1　资料与方法

1.1　资料来源

资料来源于中国疾病监测信息报告系统，病例由重庆市

各医疗卫生机构按照《传染病报告管理信息系统》上报，数据由重庆市疾控中心负责审核。本次研究选择重庆市 2006 年 1 月至 2015 年 12 月的乙脑按月报告发病例数进行时间序列分析。

1.2 分析方法

ARIMA 模型建模采用 R3.2.2 软件进行处理分析。建模过程分 3 个步骤。

1.2.1 模型识别 ARIMA 模型分析需要建立在数据序列平稳的基础上，“平稳”需满足均数和方差不随时间变化，自相关系数(autocorrelation function，ACF)只与时间间隔有关而与所处时间无关等条件。对不满足条件的序列需要进行差分平稳化，方法是运用“diff”函数，使之成为零均值、无明显变化趋势的平稳序列；对满足条件的平稳序列，分别利用“acf”和“pacf” 函数观察序列的自相关系数和偏自相关系数(partial autocorrelation function，PACF)，并根据分析结果判断数据序列适合的模型。若自相关函数拖尾且偏相关函数截尾，则数据序列适合 AR 模型；若自相关函数截尾且偏相关函数拖尾，则数据序列适合 MA 模型；若自相关函数和偏相关函数均拖尾，则数据序列适合 ARIMA 模型。ARIMA 乘积季节模型一般表示为 ARIMA(p,d,q)×(P,D,Q)s，其中 p 为非季节自回归阶数，d 为非季节差分阶数，q 为非季节移动平均阶数，P 为季节自回归阶数，D 为季节差分阶数，Q 为季节移动平均阶数，s 为季节长度。当 P=D=Q=0 时，该模型为 ARIMA 模型。P、Q 取 0、1、2 由低阶到高阶试验并根据模型拟合优度确定；s 根据疾病相关知识判断。

1.2.2 参数估计和模型检验分析 利用非线性最小二乘法或最大似然法(maximum likelihood estimation，MLE)，根据阶值计算自回归和移动平均系数。模型诊断运用“Box.test”函数计算 Box-Ljung 统计量。若统计量差异无统计学意义则认为残差为白噪声序列；若残差不能认为是白噪声则需改进模型。在残差满足白噪声的情况下，依据赤池信息量准则(Akaike's information criterion，AIC)、许瓦兹贝叶斯准则(Schwarz Bayesian criterion，SBC)和模型检验分析等筛选最优模型，AIC 和 SBC 均较小则模型拟合度越高。

1.2.3 模型预测 用筛选出的最优模型建立 2006 年 1 月至 2015 年 12 月重庆市流行性乙型脑炎发病数模型，运用“forecast.Arima”函数模型预测 2016 至 2017 年发病数并绘制预测图表。

1.3 统计分析

研究采用 Excel 2007 软件完成数据录入，用 R3.2.2 中的“tseries”和“forecast”软件包进行数据的分析处理、模型构建和模型预测。利用 2006 年 1 月到 2015 年 6 月乙脑报告病例数用于模型拟合，用 2015 年 7 至 12 月乙脑报告病例数用于模型预测效果的评价，通过实际报告病例数和拟合病例数计算各月的绝对误差、相对误差及平均绝对百分比误差来评价模型的预测效果，检验水准 $\alpha=0.05$。

2 结 果

2.1 重庆市乙脑报告病例时间分布

2006 年 1 月至 2015 年 12 月共计报告乙脑病例 2 121 例，年平均报告 212.1 例，月平均报告 17.68 例。报告病例数最多的年份是 2006 年，共 704 例；最多的月份是 2006 年 7 月，共报告 428 例。由时间序列可以看出重庆市乙脑分月报告病例数存在明显周期性，每年的报告病例高峰均出现在夏季，呈规律性季节变化，2006 至 2015 年乙脑报告病例数呈整体递减的趋势，如图 1 所示。

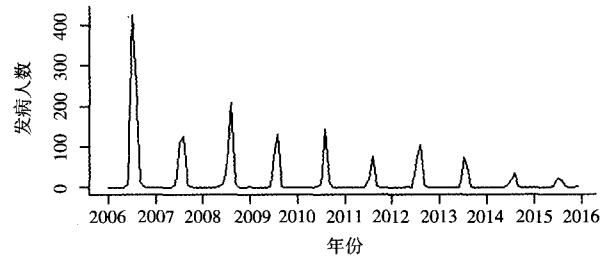


图 1 重庆市 2006 至 2015 年乙脑报告病例数时间序列(按年)

Fig.1 The reported JE cases by year from 2006 to 2015, Chongqing

2.2 建立 ARIMA 乘积预测模型

2.2.1 模型识别 应用“decompose”语句将模型按照总体趋势、季节趋势和随机误差分解，如图 2 所示。结合图 1、图 2 及乙脑的背景资料可知，本次分析的原始时间序列不满足平稳性的要求，存在以 1 年(12 个月)为 1 个周期的季节性波动，序列在延迟 12、24、36 时也具有明显的周期性，用“diff”语句对原序列进行一阶十二步季节差分消除季节的影响，得到差分后的随机序列图，如图 3 所示。其均值在 0 上下波动，可认为差分后的序列具备平稳条件，基本符合 ARIMA 乘积季节模型的平稳性要求。对差分后的数据序列分别用“acf”和“pacf”语句进行自相关和偏自相关分析，如图 4 所示。由图可知，ACF 和 PACF 均在一阶后截尾，因此初步确定 s=12，选取乘积季节 ARIMA$(p,0,q)\times(P,1,Q)_{12}$ 模型，p=1 或 0，q=1 或 0。相关研究[10]提示：P 和 Q 超过 2 阶的情况很少见，所以本研究选择对 P 和 Q 分别取 0、1、2 由低阶到高阶逐个试验，根据模型参数估计、AIC、SBC、残差序列相关系数及检验情况进行综合判断，筛选最优模型。

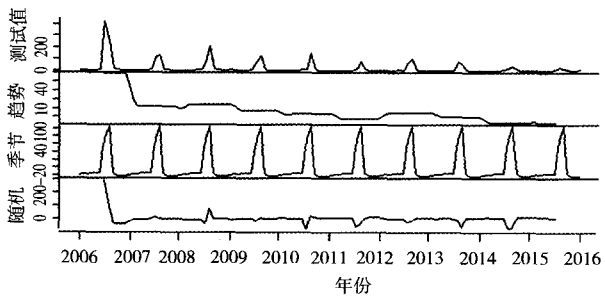


图 2 重庆市 2006 至 2015 年乙脑报告病例数分解时间序列(按年)

Fig.2 The reported decomposition of JE cases by year from 2006 to 2015, Chongqing

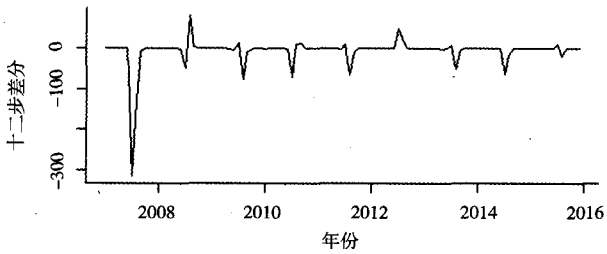


图 3 重庆市 2006 至 2015 年乙脑报告病例数差分后的时间序列(按年)

Fig.3 The reported JE cases after the seasonal difference by year from 2006 to 2015, Chongqing

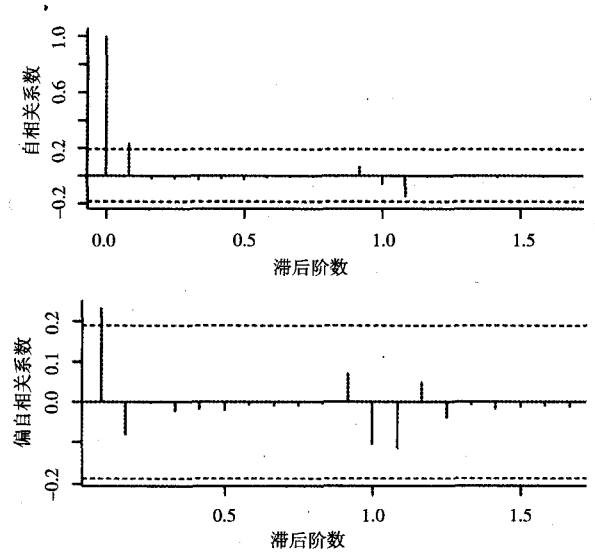


图 4 差分序列后的自相关和偏自相关系数

Fig.4 ACF and PACF after the seasonal difference

2.2.2 参数估计和模型检验分析 选择 4 个拟合的备选模型进行最优模型筛选，各参数的估计值和检验效果见表 1。评估备选模型的优劣和对时间序列在信息上的提取程度主要采用残差白噪声检验和拟合优度检验。运用“Box.test”语句计算 Box-Ljung 统计量，计算结果显示 4 个备选模型的残差均为白噪声。根据 AIC、SBC 值可知，模型 2 拟合效果最佳。综合以上结果考虑，选择 ARIMA$(0,0,1)\times(1,1,2)_{12}$ 为拟合及预测的模型。模型 2 的残差值如图 5 所示，ACF 和 PACF 分析结果如图 6 所示，差异无统计学意义（P=0.945），模型拟合效果较好。

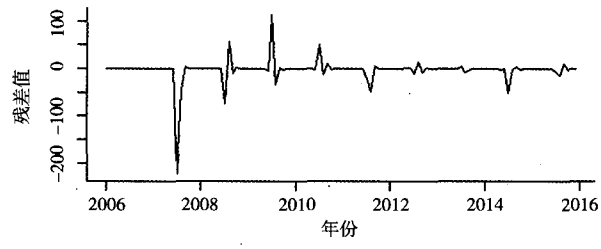


图 5 ARIMA$(0,0,1)\ (1,1,2)_{12}$ 残差值

Fig.5 The residual value of ARIMA$(0,0,1)\ (1,1,2)_{12}$

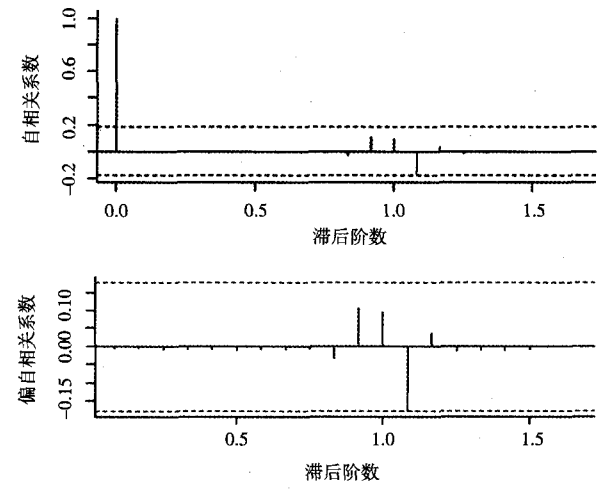


图 6 ARIMA$(0,0,1)\ (1,1,2)_{12}$ 残差的自相关和偏自相关系数

Fig.6 ACF and PACF of ARIMA$(0,0,1)\ (1,1,2)_{12}$ residual

2.2.3 模型评价 利用 ARIMA$(0,0,1)\times(1,1,2)_{12}$ 模型对重庆市 2015 年 7 至 12 月乙脑报告病例数进行预测，该模型相对误差范围位 0.0086~0.2100，平均相对误差为 0.12，平均绝对百分比误差为 7.81%，说明 ARIMA$(0,0,1)\times(1,1,2)_{12}$ 模型具有较好的预测效果，见表 2。

表 1 4 种备选模型各参数值估计及检验

Tab.1 Estimation and testing of Parameter Values of 4 alternative models

参数	ARIMA$(0,0,1)\ (0,1,2)_{12}$		ARIMA$(0,0,1)\ (1,1,2)_{12}$		ARIMA$(1,0,0)\ (0,1,2)_{12}$		ARIMA$(1,0,0)\ (1,1,2)_{12}$	
	参数估计值	标准误	参数估计值	标准误	参数估计值	标准误	参数估计值	标准误
AR1	–	–	–	–	0.181	0.095	0.196	0.095
MA1	0.186	0.094	0.204	0.095	–	–	–	–
SAR1	–	–	0.323	0.204	–	–	0.320	0.205
SAR2	–	–	–	–	–	–	–	–
SMA1	-0.500	0.085	-0.657	0.141	-0.502	0.084	-0.657	0.141
SMA2	1.000	0.189	0.999	0.190	1.000	0.187	1.000	0.189
AIC	1 076.770	–	1 076.410	–	1 076.860	–	1 076.560	–
SBC	1 107.595	–	1 095.379	–	1 117.960	–	1 101.490	–
残差序列 L-B 检验 P 值	0.880	–	0.945	–	0.929	–	0.993	–

表 2 重庆市 2015 年 7 至 12 月乙脑按月报告病例数预测结果

Tab.2 Predictive results of JE reports by month from July to December in 2015, Chongqing

月份	预测值	实际值	绝对误差	相对误差	绝对百分比误差(%)
7	20.82	21	0.18	0.01	0.86
8	29.93	26	3.93	0.15	15.00
9	2.37	3	0.63	0.21	21.00
10	1.10	1	0.10	0.10	10.00
11	0.03	0	0.03	-	-
12	-0.17	0	0.17	-	-

2.2.4 模型预测 运用 R 软件 “forecast” 函数包中的“plot.forecast”语句，选择 ARIMA$(0,0,1)\times(1,1,2)_{12}$ 模型对 2006 至 2015 年重庆市报告乙脑病例数进行拟合并预测 2016 至 2017 年乙脑发病例数，绘制的模型拟合预测序列如图 7 所示。预测 2016 年乙脑发病例数为 35 例，2017 年为 32 例，提示近两年重庆市乙脑发病水平较 2015 年略减少，发病高峰仍在夏季 7 至 8 月。

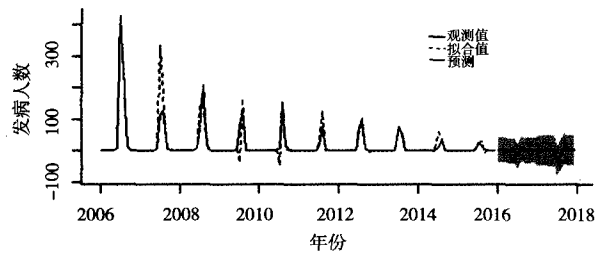


图 7 重庆 2006 至 2017 年乙脑报告病例数、拟合及预测结果(按年)

Fig.7 Reported fitted and predicted JE cases by year from 2006 to 2017, Chongqing

3 讨 论

重庆市自 2008 年开始将乙脑纳入 EPI 后报告病例数显著降低，因此对发病人数的预测对今后防治措施的调整或延续具有积极的指导意义。目前，利用数学模型进行统计预测的方法应用广泛[11]，好的模型预测有利于疾病的预警和控制。对疾病的预测目前常用且预测效果较好的方法有灰色模型[12-13]、Markov 模型[14]及 ARIMA 预测模型[15]等。乙脑作为一种蚊媒传染病，其影响因素复杂，传统的分析预测方法很难达到理想的效果，而本次研究采用的 ARIMA 乘积季节模型仅以时间作为影响因素，不考虑趋势因素、季节因素等，在长期的实际状况下达到平衡状态时作出预测[16-17]；在克服了因果回归分析法中预测对象数据资料不全问题的基础上，利用事物发展的延续性原理建立预测模型，其短期外推预测的精确度比一般线性回归模型高；此外 ARIMA 模型在一定程度上排除了人们的主观性，采用更为科学和严谨的统计手段进行疾病预测，具有重要的应用价值[18]。

从本次研究拟合和预测结果看，由于季节因素对乙脑的流行起决定性作用，重庆市乙脑的流行也出现显著的季节性，该特点在最优模型中也有很好的体现。2016 至 2017 年重庆市乙脑病例数在 2015 年的基础上有所下降，这符合重庆市乙脑发病的长期趋势，也与实际报告病例情况比较吻合（2016 年乙脑报告病例数 38 例），侧面印证了 ARIMA 模型应用于乙脑预测的可行性。预测运用的 R 软件具有开源性和强大的统计、绘图功能[19]，统计软件包由不同领域的专家共享，使用方便且功能齐全。本次预测时还发现，模型在拟合和预测时出现了部分负值，这与 ARIMA 模型本身主要从数理上反应统计规律和病例数呈减少趋势有关，即如果数据本身有递减的趋势，而且差分后的波动幅度相对不大，那么预测可能就会出现负值。在预测时，由于 ARIMA 本身不具有强加约束的处理机制，出现负值只能将其视为 0。因此，在疾病的预测中还需考虑其他因素对预测结果的影响，并且一次分析所得的模型并不一定始终能作为最好的预测依据[20]，及时的修正或重新拟合有时是必要的。总体来看，疾病预测的意义在于根据预测结果采取相应的预防控制措施，而 ARIMA 模型在疾病短期预测上有其优势。但任何模型都有其使用限制，因此多方面考虑疾病的影响因素，正确把握 ARIMA 模型利用的数据长度，同时延伸空间和时间数据能更好地提高模型预测精度及适用范围[21]。

参 考 文 献

[1] 欧柱宁. 2003-2012 年广西钦州市流行性乙型脑炎疫情分析[J]. 中国临床新医学，2013，6(9)：872-875.

[2] Campbell GL, Hills SL, Fischer M, et al. Estimated global incidence of Japanese encephalitis: a systematic review[J]. Bull World Health Organ, 2011, 89(10): 766-774.

[3] 杨绍基，任 红，李兰娟，等. 传染病学[M]. 7 版. 北京：人民卫生出版社，2008：92-98.

[4] 尹遵栋，罗会明，李艺星，等. 时间序列分析（自回归求和移动平均模型）在流行性乙型脑炎预测中的应用[J]. 中国疫苗和免疫，2010，16(5)：457-461.

临床研究

DOI:10.13406/j.cnki.cyxb.001392

重庆市九龙坡区学龄儿童贫血与体格生长的关系调查

穆丽娜，杨　亭，梁小华，陈　立，陈　洁，刘友学，李廷玉
（重庆医科大学附属儿童医院儿童保健科、儿童发育疾病研究教育部重点实验室、儿童发育重大疾病国家国际科技合作基地、儿科学重庆市重点实验室，重庆　400014）

【摘　要】目的：探讨学龄期儿童贫血与体格生长的关系。**方法**：在重庆市九龙坡区采用多阶段分层整群随机抽样的方法，对 5 所全日制公立小学的 6 260 名学龄儿童进行体格测量及血生化指标检查。统计分析贫血组及非贫血组间常用体格生长指标（体质量、身高）和与贫血相关的血常规指标，如红细胞平均容积（mean corpuscular volume，MCV）、红细胞平均血红蛋白（mean corpuscular hemoglobin，MCH）、红细胞平均血红蛋白浓度（mean corpuscular hemoglobin concentration，MCHC）和红细胞分布宽度（red blood cell distribution width，RDW）的差异及相关性。**结果**：4 730 例有效样本中，共检出 1 140 例贫血儿童，贫血的总检出率为 24.1%，女童贫血的检出率显著高于男童（男童检出率为 22.6%，女童检出率为 25.7%，χ^2=6.100，P=0.010）。各年龄组（6 岁~、7 岁~、8 岁~、9 岁~、10 岁~、11 岁~、12 岁~）贫血检出率分别为 29.1%、29.1%、27.6%、23.7%、20.9%、17.4%、10.5%。不同年龄组间贫血检出率存在显著差异（χ^2=75.800，P=0.000）。贫血组与非贫血组间的体格生长指标、血常规指标均存在显著性差异（P<0.05）。贫血对儿童的体质量及身高均具有显著影响（分别为 F=124.400，P=0.000；F=93.650，P=0.000）。血红蛋白（Hb）浓度与体质量、身高、体质指数（body mass index，BMI）、MCV、MCH、MCHC 均具有显著相关性（相关系数分别为 0.233、0.221、0.174、0.277、0.427、0.319，P 值均为 0.000）。**结论**：重庆市九龙坡区学龄儿童贫血的发生率较高；贫血的发生与年龄、性别、体质量和身高均存在相关关系；贫血对学龄期儿童的体质量及身高均具有显著影响。提示重庆市九龙坡区学龄儿童贫血情况需高度重视，应该加强对该类儿童家长及儿童本身的营养宣教，提高其对营养的关注度，合理膳食。

【关键词】贫血；体格生长指标；营养
【中图分类号】R72　　**【文献标志码】**A　　**【收稿日期】**2016-04-27

作者介绍：穆丽娜，Email：mulina11@163.com，
研究方向：儿童营养。
通信作者：李廷玉，Email：tyli@vip.sina.com。
优先出版：http://kns.cnki.net/kcms/detail/50.1046.R.20170912.1031.004.html
（2017-09-12）

[5] 陆　波，闵思韬，闵红星，等. 应用 ARIMA 模型预测麻疹发病率的可行性研究[J]. 中国卫生统计，2015，3(1)：106-107.
[6] 黄春萍，刘仕俊，丁　华，等. 杭州市麻疹发病数的 ARIMA 模型预测[J]. 现代预防医学，2014，41(11)：1935-1937.
[7] 刘继恒，白春林，孙要武，等. 应用 ARIMA 模型预测宜昌市手足口病的发病趋势[J]. 中国热带医学，2014，14(8)：956-958.
[8] 冯　冰，张　艳，周　方. 武汉市江汉区细菌性痢疾发病率的 ARIMA 模型预测应用[J]. 重庆医学，2015，44(29)：4123-4124.
[9] 孟　蕾，王新华，白亚娜，等. 甘肃省哨点医院流感样病例 ARIMA 模型预测[J]. 中国公共卫生，2014，30(2)：228-230.
[10] 张　蔚，张彦琦，杨　旭. 时间序列资料 ARIMA 季节乘积模型及其应用[J]. 第三军医大学学报，2002，24(8)：955-957.
[11] 孙振球. 医学统计学[M]. 北京：人民卫生出版社，2005：451-492.
[12] 解华山. 灰色模型 GM(1，1)模型预测传染病时样本量的应用探讨[J]. 公共卫生与预防医学，2013，24(5)：51-53.
[13] 刘　洁，何钦成. 我国流行性乙型脑炎发病的聚类分析及其未来发病趋势预测[J]. 中国媒介生物学及控制杂志，2013，24(2)：152-154，160.
[14] 樊雯婧，陆　群，邹立巍，等. Markov 模型在合肥市疟疾发病趋势预测中的应用[J]. 中国热带医学，2013，13(7)：819-821.
[15] 夏　菁，张华勋，林　文，等. ARIMA 模型在疟疾发病率预测中的应用[J]. 中国血吸虫病防治杂志，2016，28(2)：135-140.
[16] 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.
[17] 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.
[18] 樊雯婧，陆　群，邹立巍，等. ARIMA 模型在合肥市疟疾发病预测中的应用[J]. 安徽医科大学学报，2013，48(3)：252-256.
[19] 易延华，胡桂洁. 应用 R 软件和 Epicalc 程序包进行医学 Logistic 回归分析[J]. 中国病案，2015，16(3)：33-36.
[20] 温　亮，徐德忠，林明和，等. 应用时间序列模型预测疟区疟疾发病率[J]. 第四军医大学学报，2004，25(6)：507-510.
[21] 周水森，黄　芳，沈毓祖. ARIMA 模型在疟疾发病预测中的应用[J]. 中国病原生物学杂志，2007，2(4)：284-286.

（责任编辑：冉明会）