

• 流行病学研究 •

基于配对检验的ARIMA模型在我国甲肝发病数预测中的应用

丁 勇¹, 张蓓蓓^{1*}, 吴 静²¹南京医科大学康达学院, 江苏 连云港 222000; ²南京医科大学生物医学工程与信息学院, 江苏 南京 211166

[摘要] 目的: 探讨基于配对检验的求和自回归移动平均(autoregressive integrated moving average, ARIMA)模型在我国甲肝发病预测中的应用, 提出时间序列模型预测效果评价的新思路与方法。方法: 根据2004年1月—2021年12月我国甲肝传染病月发病数建立ARIMA模型, 对2022年1—8月的甲肝月发病数进行预测, 通过配对 t 检验和误差分析评估该模型的预测效果。结果: 配对 t 检验结果显示, ARIMA(1, 1, 0)(0, 1, 1)₁₂模型预测的甲肝月发病数与实际月发病数差异无统计学意义($P > 0.05$), 说明模型有较好的预测能力, 预测结果的相对误差平均值为3.86%, 标准差为3.25%。结论: ARIMA乘积季节模型能够较准确地预测我国甲肝的发病趋势; 配对检验为时间序列模型预测效果的评价提供了客观评价依据, 较好地解决了时间序列模型预测效果的评价问题。

[关键词] 配对检验; 甲型肝炎; ARIMA乘积季节模型; 预测

[中图分类号] R512.6

[文献标志码] A

[文章编号] 1007-4368(2024)10-1456-06

doi: 10.7655/NYDXBNSN240080

Application of ARIMA model based on paired test in the prediction of hepatitis A incidence in China

DING Yong¹, ZHANG Beibei^{1*}, WU Jing²¹Kangda College, Nanjing Medical University, Lianyungang 222000; ²School of Biomedical Engineering and Information, Nanjing Medical University, Nanjing 211166, China

[Abstract] **Objective:** To explore the application of autoregressive integrated moving average (ARIMA) model based on paired test in predicting the incidence of hepatitis A in China, and put forward a new idea and method for evaluating the prediction effect of time series model. **Methods:** An ARIMA model was established for the monthly incidence of hepatitis A infectious diseases in China from January 2004 to December 2021, and the monthly incidence of hepatitis A infectious diseases from January to August 2022 was predicted. The prediction effect of the model was evaluated by paired t -test and error analysis. **Results:** The results of paired t -test showed that there was no significant difference between the monthly incidence of hepatitis A predicted by ARIMA(1, 1, 0)(0, 1, 1)₁₂ model and the actual monthly incidence of hepatitis A ($P > 0.05$), indicating that the model had good prediction ability, and the mean relative error and standard deviation of the prediction results were 3.86% and 3.25%. **Conclusion:** ARIMA product season model can accurately predict the incidence trend of hepatitis A in China. The paired test provides an objective basis for evaluating the prediction effect of time series model, and solves the problem of evaluating the prediction effect of time series model well.

[Key words] paired test; hepatitis A; multiple seasonal ARIMA model; prediction

[J Nanjing Med Univ, 2024, 44(10): 1456-1461]

[基金项目] 国家自然科学基金(61901225); 江苏省高校自然科学研究(19KJD330001); 江苏高校哲学社会科学研究(2022SJYB1868); 江苏省大学生创新创业训练计划(202213980011Y); 南京医科大学康达学院第一期青年教师科研导师项目(KD2022KYDS020); 南京医科大学康达学院第二期品牌专业建设工程(JX206000302); 南京医科大学康达学院医学信息模拟及预测科研团队(KD2022KYCXTD003)

*通信作者(Corresponding author), E-mail: bbzhang@njmu.edu.cn

甲型病毒性肝炎(简称甲肝)是由甲肝病毒(hepatitis A virus, HAV)引起的,以肝脏损伤为特征的急性传染病,因摄入受到污染的食物、水或直接接触感染者而致病,粪-口途径为甲肝的主要传播方式^[1]。全球范围内甲肝发病呈散发趋势^[2],2017年,全球新增病例1.7亿例。我国是甲肝的中-高流行区,1988年的上海甲肝大暴发,疫情造成罹患人数达31万之多,带来了极大的社会恐慌和财产损失^[3]。因此,探索甲肝发病流行特征,建立一个高精度的预测模型,提出有效的模型预测评估方法,对我国甲肝传染病防控具有重要意义。

目前,求和自回归移动平均(autoregressive integrated moving average, ARIMA)模型在传染病建模与预测研究中得到广泛应用。模型的拟合效果常采用拟合优度 R^2 、正态化贝叶斯信息准则(Bayesian information criterion, BIC)、模型参数检验等相关指标和方法进行评价;预测效果一般采用误差分析进行判断,例如绝对误差、相对误差等,用以确定预测值与实际值的差异,但这种预测效果判断缺乏客观、统一的评价标准,在一定程度上影响了ARIMA预测模型的应用价值。本研究以我国的甲肝发病数为例,提出了将配对检验应用于ARIMA模型预测效果评价的方法,旨在解决时间序列模型预测效果的评价问题,为完善时间序列模型的应用提供新思路和方法。

1 资料和方法

1.1 资料

甲肝的月发病数资料来源于中国疾病预防控制中心网站(<http://www.nhc.gov.cn/jkj/>)2004年1月—2022年8月的全国法定传染病疫情报告。

1.2 方法

本文使用统计分析软件SPSS27.0进行数据整理和分析。

1.2.1 模型建立

ARIMA模型是基于时间序列数据自身的特性,通过对数据的自回归、移动平均和差分等变换,来建立描述数据特征的预测模型,并应用该模型预测未来短期内的数据变化规律^[4]。季节ARIMA模型适用于由于季节性变化(包括季度、月度等)或其他一些因素引起的周期性变化的时间序列。季节ARIMA模型可表示为 $ARIMA(p, d, q)(P, D, Q)_s$,其中 s 为季节的周期长度, d 为差分阶数, D 为季节差分阶数, p 为自回归阶数, P 为季节自回归阶数, q 为

移动平均阶数, Q 为季节移动平均阶数。

对我国2004年1月—2021年12月的甲肝月发病数建立ARIMA时间序列模型,并应用该模型对2022年1—8月的发病数进行预测。

1.2.2 预测能力检验

一个优良的时间序列模型应具有较好的预测能力,本研究提出用配对检验的方法^[5]对模型的预测能力进行统计检验。

当同一个月发病数的模型预测值与实际值的差值服从正态分布时,使用配对 t 检验检验模型的预测能力;当差值不服从正态分布时,使用配对样本非参数检验的Wilcoxon符号秩检验进行预测能力检验。对给定的显著性水平($\alpha=0.05$),当接受原假设时,认为模型的预测结果与实际结果无显著性差异,模型具有预测能力,有实际应用价值;否则模型的预测能力较差,没有实际应用价值。

2 结 果

2.1 流行特征分析

我国2004年1月—2021年12月的甲肝发病人数时间序列如图1所示,我国甲肝的发病人数整体呈现下降趋势,由此可见,近年来我国甲肝的防治工作卓有成效。甲肝发病具有季节性周期特征(周期 $s=12$ 个月),通过时间序列的季节性分解^[6],发病数季节因子如图2所示,季节因子 <0 表示季节水平低于平均值,季节因子 >0 表示高于平均值。通过季节因子可知我国甲肝发病的流行特性周期性特征为:1—2月发病人数较少,其中2月发病人数最少;7—9月发病人数较多,是防控工作的重点,其中8月发病人数最多,这可能与夏天气温高、雨水多,暑期放假人员流动量大,病毒易于扩散有关。

2.2 建立预测模型

季节ARIMA模型建模要求时间序列是平稳的,而图1表明甲肝数据序列不平稳,通过对发病原始时间序列进行自然对数转化和一阶差分($d=1$)、一阶季节差分($D=1$)的变换,时间序列达到平稳(图3),序列平稳性的扩展的迪基-富勒(augmented Dickey-Fuller, ADF)检验^[7]显著性 $P < 0.01$ 。

根据甲肝发病时间序列的周期 $s=12$ 及平稳化差分处理($d=1, D=1$),模型初定为 $ARIMA(p, 1, q)(P, 1, Q)_{12}$ 。由平稳化处理后序列的自相关、偏自相关图的截尾情况,选取 $p=1$ 和 $q=1$;根据已有研究成果可知^[8-9], P, Q 取值超过2阶的情形一般比较少见,本研究分别取 P, Q 为0、1、2进行计算,从而有

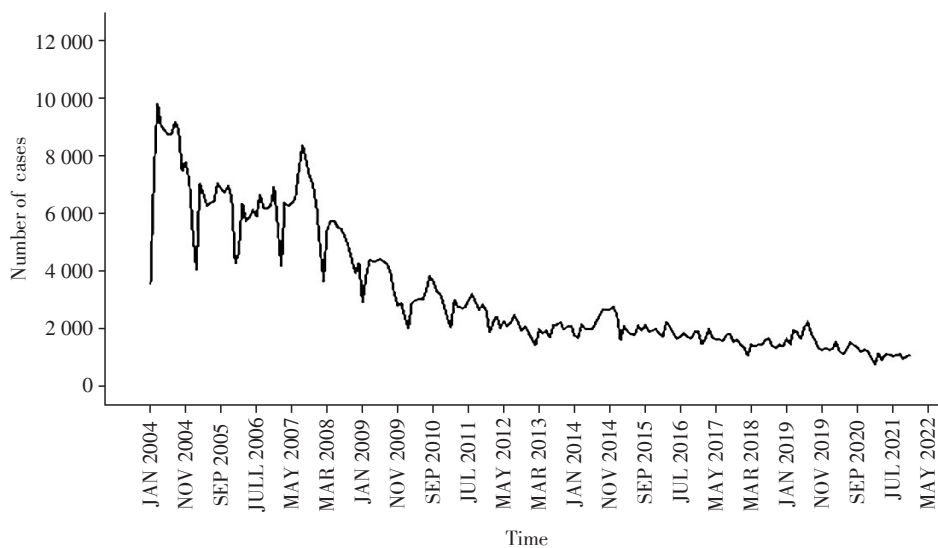


图1 我国甲肝发病人数时间序列图

Figure 1 Time series of incidence of hepatitis A in China

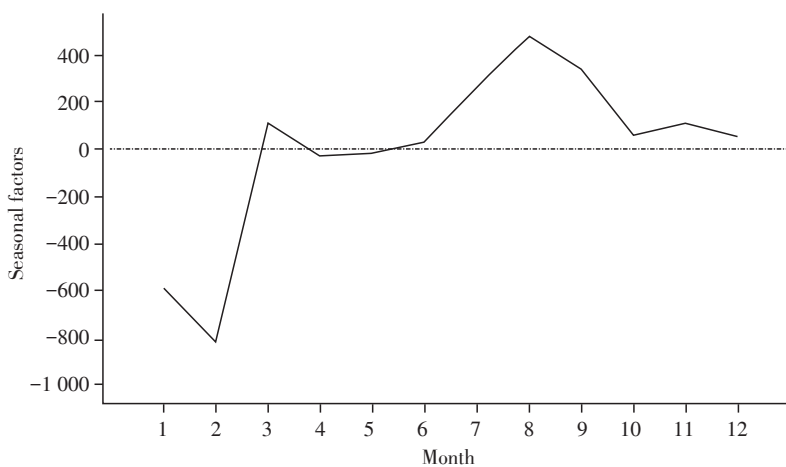


图2 甲肝发病数的季节因子

Figure 2 Seasonal factors of incidence of hepatitis A

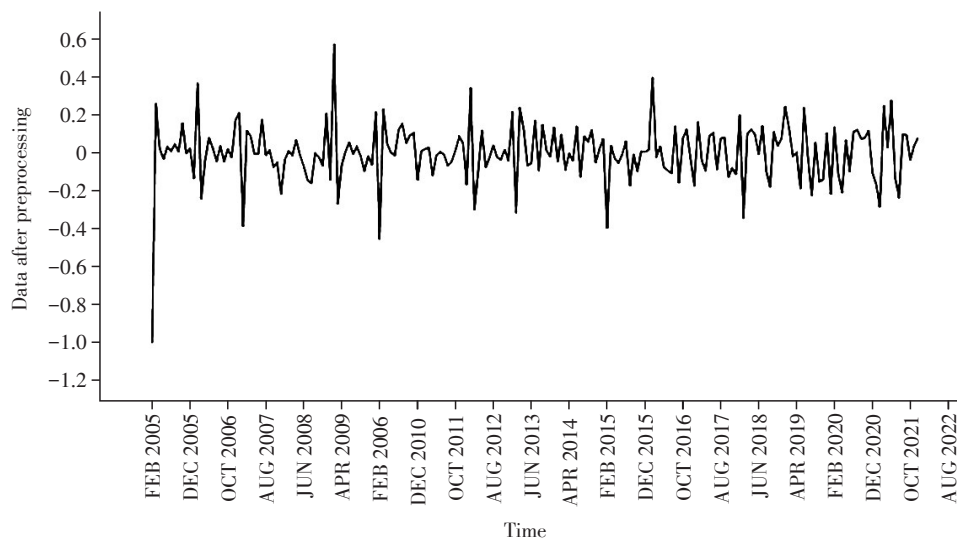


图3 平稳的时间序列图

Figure 3 Stationary time series

3²=9种备选模型。根据杨-博克斯显著性检验、模型系数显著性检验,以及模型拟合残差的自相关、偏自相关函数图(图4)判断残差符合白噪声序列的要求,选出模型ARIMA(1,1,0)(0,1,1)₁₂。表1说明模型ARIMA(1,1,0)(0,1,1)₁₂的系数均通过 t 检验($P < 0.05$);拟合优数 R^2 和BIC值表现良好。从图5模型ARIMA(1,1,0)(0,1,1)₁₂的拟合曲线来看,拟合效果良好,拟合值与实测值比较接近,能够较好地动态反映甲肝时间序列的波动规律和季节特性。

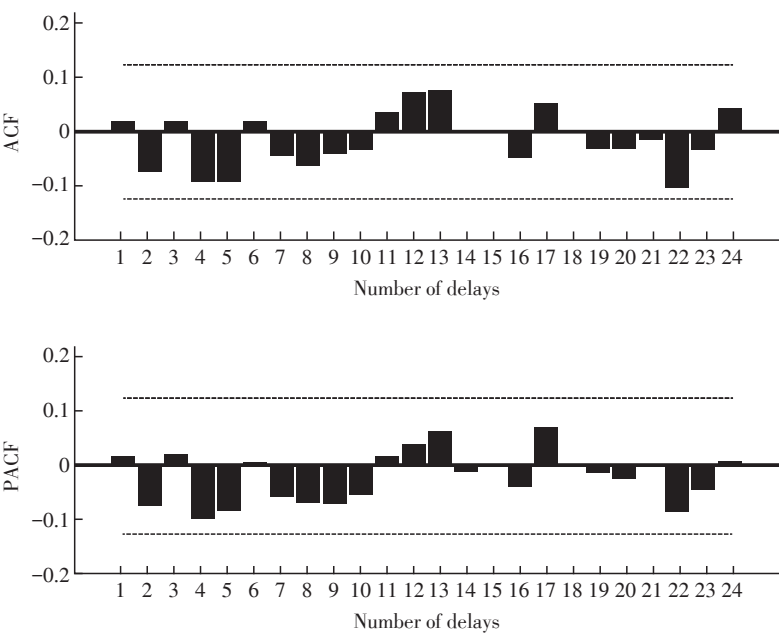


图4 残差自相关函数(ACF)图和偏自相关函数(PACF)图
Figure 4 Residual autocorrelation function(ACF) and Partial autocorrelation function(PACF)

表1 ARIMA(1,1,0)(0,1,1)₁₂模型参数估计检验及拟合结果统计

Table 1 Statistics of parameter estimation test and fitting results of ARIMA(1,1,0)(0,1,1) ₁₂ model						
Variable	AR(1)	SMA(12)	Constant	Box's Q test statistic	R ²	BIC
Coefficient	-0.284	0.799	< 0.001	—	0.891	12.936
t	-4.481	13.443	-0.278	—		
P	< 0.001	< 0.001	0.781	0.664		

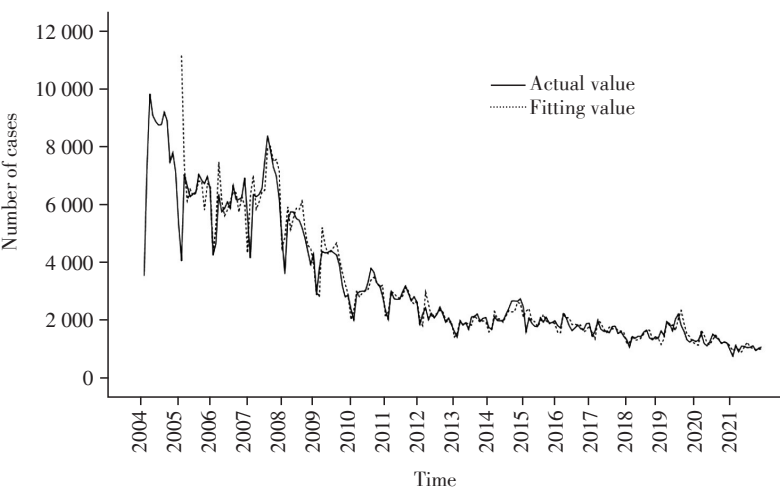


图5 ARIMA(1,1,0)(0,1,1)₁₂的拟合曲线图
Figure 5 Model fitting of ARIMA(1,1,0)(0,1,1)₁₂

表2 ARIMA(1,1,0)(0,1,1)₁₂模型预测结果
Table 2 Prediction results of ARIMA(1,1,0)(0,1,1)₁₂ model

Variable	Month								Mean relative error	Relative error standard deviation
	1	2	3	4	5	6	7	8		
Actual value	956	830	1 050	896	960	1 038	1 069	1 116	—	—
Predictive value	934	846	1 125	988	993	1 029	1 108	1 102	—	—
Difference	-22	16	75	92	33	-9	39	-14	—	—
Relative error(%)	2.30	1.93	7.14	10.27	3.44	0.87	3.65	1.25	3.86	3.25

2.3 模型预测效果检验

作为实际应用,通过数据拟合建立模型是第一步,模型拟合得好,仅说明模型对已有数据有较好的适应性,但对未来数据的预测能力还需进一步检验。为了进一步检验ARIMA(1,1,0)(0,1,1)₁₂模型的实际预测应用能力,利用ARIMA(1,1,0)(0,1,1)₁₂模型对2022年1—8月的甲肝发病情况进行预测,得到预测值的结果如表2所示。

根据夏皮洛-威尔克(Shapiro-Wilke)正态性检验方法,甲肝发病数实际值与模型预测值差值的正态性检验 $P=0.51$,满足正态性要求。对预测值与实际值进行配对 t 检验,结果显示预测组与实际组差值的均数和标准差分别为-26.250和41.762,标准误差的平均值为14.754,差值的95%置信区间为-61.157~8.657, $t=-1.778$, $P=0.119$,不拒绝原假设,认为预测值与实际值无显著性差异,说明模型具有较好的预测能力。在模型具有预测能力的基础上,可以进一步分析模型的预测误差。表2显示ARIMA(1,1,0)(0,1,1)₁₂模型预测的相对误差平均值为3.86%,标准差为3.25%,相对误差标准差较小,说明预测结果比较稳定;相对误差平均值较小,说明预测误差较小,相比文献预测的平均相对误差4.69%~21.69%^[10-12],说明ARIMA(1,1,0)(0,1,1)₁₂模型有较好的预测效果。

如果应用配对样本非参数检验的Wilcoxon符号秩检验,结果为 $P=0.123$,认为预测值与实际值总体无显著性差异,也可以得到相同的结论。

3 讨 论

甲肝是我国的常见、多发传染病,甲肝防控是保障公共卫生安全的重点工作之一,开展甲肝流行特征的分析 and 预测是对疾病未来发生、发展及流行趋势认知的重要手段,也是制定甲肝防控策略的重要前提。

时间序列模型目前已被广泛应用于传染病发病的预测,一个好的模型,不仅模型的拟合效果好,

而且预测效果也好,才有真正的实际应用价值,因此,预测效果的评价具有重要意义。但目前模型的预测效果评价尚无客观、统一的依据,通过分析维普中文期刊服务平台2000—2023年应用ARIMA模型预测肝炎发病数的50余篇论文,发现其中有文献认为模型拟合得好,便可以进行预测^[13];有文献通过描绘发病预测值与实际值的数据点,直观观察两者数据曲线吻合程度进行评价^[14];有文献通过实际值是否位于预测值的95%置信区间进行评价^[15]。而大多数文献则通过一些误差指标进行评价,例如绝对误差(absolute error, E)、绝对误差率(absolute error rate, ER)、平均绝对误差(mean absolute error, MAE)、均方误差(mean square error, MSE)、均方根误差(root mean square error, RMSE)、平均绝对百分比误差(mean absolute percentage error, MAPE)等。这些指标虽然可以给出预测误差的大小,但没有一个统一的指标作为标准。即使使用同一个指标,比如平均相对误差,评价时依据应该是<10%,还是<15%或20%,也没有标准。以季节ARIMA模型对传染病的预测为例,文献报道,我国戊肝发病趋势的预测平均相对误差为4.69%,认为预测效果良好,能够较好地描述该时段戊肝的发病趋势^[10];湖北省甲肝发病预测的平均相对误差率为13.09%,认为预测值和实际数值吻合良好^[11];山东省肺结核发病的预测平均相对误差为21.69%,也给出了模型能够较好地预测发病趋势的结论^[12]。因此对预测效果的评价有必要给出一种方法与标准作为客观评价的判断依据。本研究提出了应用配对检验解决这一问题的方法:当预测值与实际值无显著性差异时,说明模型有较好的预测能力,然后再结合一些误差指标进一步说明预测误差的大小和稳定性。ARIMA模型预测效果仅用误差指标来评价是不全面的,预测效果的评价应包含两部分,预测能力评价和预测误差分析,如果模型不具备预测能力,分析预测误差大小意义不大。本研究以我国常见多发传染病甲肝为例,建立ARIMA预测模型,对

2022年1—8月甲肝月发病数进行预测,通过正态性检验及配对 t 检验表明模型预测值与实际值无显著性差异,说明模型有预测能力,再结合预测结果的误差分析,说明模型预测精度较高,稳定性较好,模型预测有一定实用价值。

传染病预测模型是将流行病学的有关理论与数学、统计学、计算机科学、系统科学等学科理论进行结合,通过建立数学模型来模拟传染病的流行传播过程,从而实现对疫情发展趋势的预测、对不同防控措施效果的评估等目的^[16]。目前最常用的模型有ARIMA模型,其他还有微分灰色预测模型、Markov模型、通径分析模型等,主要依赖数据自身特性建模^[17-18]。季节ARIMA模型法结合了自回归和移动平均方法的优点,对各类型时间序列数据具有较强的适用性和灵活性,可以较好地描述传染病流行的趋势性、季节性和非线性特征,得到了较多应用。本研究以ARIMA模型为例,说明了配对检验在模型预测效果中的应用,由于配对检验只对模型的预测结果与实际值进行比较,与模型无关,因此本研究的方法可推广到其他各种预测模型,具有应用的广泛性。

基于配对检验的评价方法,完善了时间序列模型的预测评价,提供了时间序列模型预测效果评价的新思路。对于时间序列预测模型的构建,需要疾病发病数据的及时补充,才能即时体现预测的时效,一般而言,短期预测效果要好于长期预测。当预测值与实际值的差值服从正态分布时,应用参数检验的配对 t 检验,此时检验效能高;当预测值与实际值的差值不服从正态分布时,应用非参数的配对检验,检验效能低于参数检验方法。

[参考文献]

- [1] GUPTA V. Hepatitis A: signs and symptoms, transmission, diagnosis and treatment[J]. J Hematol Thromb Dis, 2021,9(9):1-2
- [2] 朱佳佳,胡登利,洪秀琴,等.基于时空大数据的甲型肝炎发病率分布特征分析及预测模型[J]. 中华疾病控制杂志,2018,22(11):1144-1147
- [3] 蔡文静. 1988年上海甲肝疫情研究[D]. 南昌:江西师范大学,2022
- [4] 马一鸣,丁 勇. 我国2012—2021年4种肝炎流行趋势

的时间序列分析和预测[J]. 南京医科大学学报(自然科学版),2024,44(1):72-79

- [5] 颜 艳,王 彤. 医学统计学[M]. 5版. 北京:人民卫生出版社,2020:40-41,125-128
- [6] 邓维斌,付 强,周玉敏,等. SPSS统计分析实用教程[M]. 3版. 北京:电子工业出版社,2023:279-281
- [7] 陈 刚,郭 鹏,皮 鹤,等. 基于单位根检验和ARMA模型的桥墩位移稳定性时间序列分析[J]. 武汉工程大学学报,2023,45(5):586-590
- [8] 张蓓蓓,彭献镇,王建明,等. 中国肺结核发病趋势的ARIMA乘积季节模型构建[J]. 江苏预防医学,2021,32(4):400-402
- [9] 游楠楠,刘 巧,李忠奇,等. 基于ARIMA模型的江苏省不同地区肺结核发病趋势的预测[J]. 南京医科大学学报(自然科学版),2020,40(6):909-914
- [10] 丁 勇,吴 静,武 丹,等. ARIMA乘积季节模型预测我国戊肝的发病趋势[J]. 南京医科大学学报(自然科学版),2020,40(11):1725-1729
- [11] 李丽娜,李 宁,胡 樱,等. 2004—2019年湖北省甲型肝炎流行特征分析及预测[J]. 中华疾病控制杂志,2020,24(10):1165-1169
- [12] 田 庆,刘永鹏,张晶晶,等. ARIMA乘积季节模型在山东省肺结核发病预测中的应用[J]. 山东大学学报(医学版),2021,59(7):112-118
- [13] 陈正利,陈 伟,许汴利. 应用ARIMA模型对河南省1991—2011年乙型肝炎发病趋势分析[J]. 中国卫生统计,2013,30(3):401-402
- [14] 吴田勇,曾 庆,于 萌,等. 2004—2012年中国丙型肝炎报告数据ARIMA模型及其趋势预测[J]. 上海交通大学学报(医学版),2014,34(5):705-709
- [15] 钱晓涵,郑 敏,姚永明,等. 贵州省2010—2022年梅毒流行特征分析及ARIMA模型预测[J]. 应用预防医学,2023,29(6):383-387
- [16] 沈世敬,崔晓鸣,曹务春. 传染病预测模型的应用及研究进展[J]. 中华医院感染学志,2023,33(16):2550-2554
- [17] 李琼芬,黄 甜,王荣华,等. 传染病疫情预测预警模型研究进展[J]. 中国公共卫生,2013,29(11):1695-1697
- [18] 张生奎,王镇德,杨 荔,等. 基于SARIMA-ERNN组合模型预测我国细菌性痢疾发病率[J]. 南京医科大学学报(自然科学版),2019,39(6):925-931

[收稿日期] 2024-01-20

(本文编辑:蒋 莉)