

• 诊治分析 •

Holt-Winters 模型与 SARIMA 模型对手足口病
月发病率预测效果的比较

陈亚琴, 陈健, 朱诚, 朱利华

【关键词】 手足口病; 预测; Holt-Winters 模型; SARIMA 模型; 时间序列分析

doi:10.3969/j.issn.1671-0800.2024.07.016

【中图分类号】 R181 【文献标志码】 A 【文章编号】 1671-0800(2024)07-0901

手足口病(hand, foot and mouth disease, HFMD)是由肠道病毒感染引起的儿童常见传染病,以发热、手足臀部或者口腔疱疹为主要症状,多发于 5 岁以下儿童,严重时可导致患儿死亡^[1-2]。2008 年 3 月,我国把 HFMD 列为丙类传染病进行管理^[3],近年仍是威胁儿童健康的重要公共卫生问题之一^[4]。Holt-Winters 指数平滑(Holt-Winters exponential smoothing, Holt-Winters)模型和季节差分自回归移动平均(seasonal auto-regressive integrated moving average, SARIMA)模型均有预测 HFMD 发病率的效能,本研究拟比种两组预测模型对 HFMD 月发病率的预测效能,现报道如下。

1 资料与方法

1.1 数据来源 本研究使用的 HFMD 月报告发病数据来源于中国疾病预防控制中心发布的“每月全国法定传染病疫情概况(不含香港、澳门特别行政区和台湾地区)”,人口数据来源于国家统计局^[5-6]发布的 2013—2023 年年末常住人口数据(包括 31 个省、自治区、直辖市和现役军人的人口,不包括居住在 31 个省、自治区、直辖市的港澳台居民和外籍人员),以此为依据测算出 2013 年 1 月至 2023 年 12 月全国 HFMD 月发病率,共计 132 个数据。采用 2013 年 1 月至 2022 年 12 月的 HFMD 月发病率数据作为训练集,建立模型。采用 2023 年 1 ~ 12 月 HFMD 月发病率数据作为测试集来评估模型的预测效能。

1.2 模型与评价方法 Holt-Winters 模型分加法模型和乘法模型两种^[7],该模型可近似表示为: $Y = level + slope \times t + s_t + irregular_t$,其中, Y 表示特定时间

点 HFMD 发病率;平滑参数 α 用于控制表示常数水平项的 $level$;斜率参数 β 用于控制表示趋势项的 $slope$;光滑参数 γ 用于控制表示 t 时刻的季节效应的 s_t , $irregular_t$ 表示 t 时刻的随机项。

SARIMA 模型是基于自回归移动平均模型(ARIMA 模型)的进一步发展^[8],特别适用于序列中同时存在趋势性和季节性的情况^[9-10]。SARIMA 模型简记为 $SARIMA(p, d, q)(P, D, Q)_s$ 。其中 p, d 和 q 分别代表自回归、差分次数和移动平均阶数, P, D 和 Q 分别代表季节性自回归、季节性差分次数和移动平均阶数, s 代表季节周期和循环长度^[11]。

利用两种模型分别预测 2023 年 1 ~ 12 月 HFMD 月发病率,并与测试集相比较,对照分析两种模型预测值的均方根误差(root mean squared error, $RMSE$)、平均绝对误差(mean absolute error, MAE)、平均绝对百分误差(mean absolute percentage error, $MAPE$)和决定系数(R^2)。

1.3 统计方法 采用 Excel2010 软件建立数据库,利用 Python3.8 软件的 statsmodels(veision0.13.5)函数建立模型并预测。 $P < 0.05$ 表示差异有统计学意义。

2 结果

2.1 HFMD 发病率的时序特征分析 2013 年 1 月至 2023 年 12 月,我国 HFMD 发病呈现季节性波动趋势,每个周期长达 12 个月,其中月发病率最高值为 2014 年 5 月(38.416/10 万),最低值为 2023 年 1 月(0.176/10 万)。除 2020 年 9 ~ 11 月发病趋势为较低单峰外,其余年度发病率呈现双高峰波动态势:4 ~ 7 月为大高峰,9 ~ 11 月为小高峰,期间发病高峰峰值有下降趋势,至 2023 年 1 月达到低谷,此后转变为显著上升的单高峰形态,见图 1。

作者单位: 315010 宁波,宁波大学附属妇女儿童医院

通信作者: 朱利华, Email: onlylihua@126.com

进一步分析序列的季节性分布特征和趋势性,发现 HFMD 发病率呈现一个整体下降后回升的趋势,且具有周期长度为 12 个月的季节变动。观察序列的随机效应,发现提取出季节效应和趋势后的残差基本稳定,见图 2。

2.2 Holt-Winters 模型构建与预测分析 采用 Holt-Winters 模型的加法和乘法模型进行构建,模型平滑参数由系统自动设定,得出两种模型的拟合参数见表 1。因乘法模型 $R^2 < 0$,表明效果较差。因此选择 Holt-Winters 加法模型预测 2023 年 1~12 月发病率,结果显示预测值均在 95%CI 内, R^2 为 0.313,提示模型拟合效果良好,见图 3。将预测值与测试集进行比较,计算得到 Holt-Winters 加法模型预测的 RMSE 为 8.866, MAE 为 6.933, MAPE 为 327.069%。

2.3 SARIMA 模型构建与预测分析 对原始数据集进行单位根检验,显示原始序列存在单位根,序列不平稳(ADF 检验: $t=1.96, P > 0.05$),见表 2。对原始序列进行 1 阶 12 步差分后序列平稳(ADF 检验: $t=4.73, P < 0.05$)。绘制 1 阶 12 步差分序列的自相关函数(ACF)及偏自相关函数(PACF)图,见图 4,示自相关系

数拖尾,并呈现季节效应,提示 SARIMA 模型适用于 HFMD 时间序列的模型构建。

利用网格搜索的方法对模型可能的参数值进行拟合,利用赤池信息量准则(akaike information criterion, AIC)来评价所有模型,最终 SARIMA(1,0,1)(0,1,1)₁₂ 模型 AIC 值最低(AIC=504.46),选为最优模型。对 SARIMA(1,0,1)(0,1,1)₁₂ 模型进行参数估计,得到模型的各个参数差异均有统计学意义(均 $P < 0.05$),见表 3。

利用构建的 SARIMA(1,0,1)(0,1,1)₁₂ 模型对 2023 年 1~12 月全国 HFMD 发病率进行预测,结果显示预测值均在 95%CI 内, R^2 为 0.440。将预测值与测试集进行比较,计算得到 RMSE 为 8.004, MAE 为 5.433, MAPE 为 159.949%,见图 5。

表 2 HFMD 时间序列差分前后 ADF 检验结果

参数	原始序列	1 阶 12 步差分序列
检验统计量	1.96	4.73
P 值	> 0.05	< 0.05
临界值(1%)	3.49	3.49
临界值(5%)	2.89	2.89
临界值(10%)	2.58	2.58

表 3 SARIMA(1,0,1)(0,1,1)₁₂ 模型参数统计分析

参数	估计值	标准误	Z 值	P 值	95%CI
ar.L1	0.7245	0.079	9.13	< 0.05	0.569 ~ 0.880
ma.L1	0.3626	0.136	2.66	< 0.05	0.069 ~ 0.629
ma.S.L12	-0.7129	0.101	7.09	< 0.05	-0.910 ~ -0.516
sigma2	11.0815	1.439	7.70	< 0.05	8.261 ~ 13.902

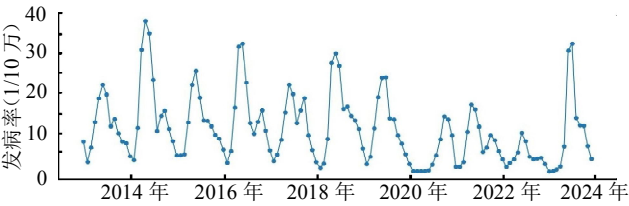


图 1 2013 年 1 月至 2023 年 12 月全国 HFMD 发病率时间序列图

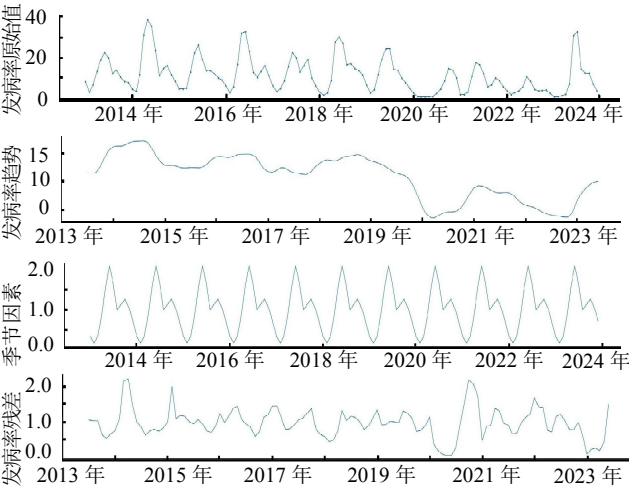


图 2 2013 年 1 月至 2023 年 12 月 HFMD 月发病率时间序列的趋势及季节分解图

表 1 Holt-Winters 模型拟合参数评价

模型	R^2 值	α 值	γ 值	β 值
Holt-Winters 乘法模型	-0.404	1.000	7.68×10^{-9}	0.009
Holt-Winters 加法模型	0.313	1.000	5.19×10^{-10}	0.000

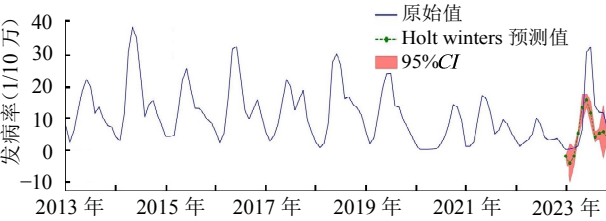


图 3 Holt-Winters 模型预测图

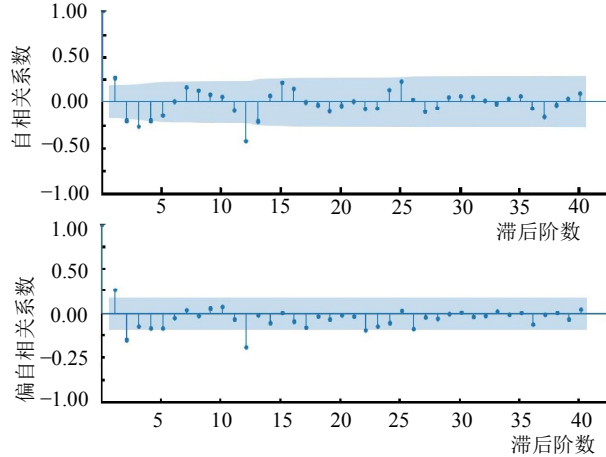


图 4 HFMD1 阶 12 步差分序列 ACF 和 PACF 图

2.4 预测效果比较 两者相比, SARIMA (1, 0, 1) (0, 1, 1)₁₂ 模型的预测值与实际值更接近, 预测值曲线趋势更拟合 2023 年实际发病趋势, 3 项误差值更小, 且 R^2 值高于 Holt-Winters 加法模型, 更接近于 1, 即 SARIMA 模型预测效果优于 Holt-Winters 模型, 见图 6。

3 讨论

本研究结果显示, Holt-Winters 加法模型与 SARIMA (1, 0, 1) (0, 1, 1)₁₂ 模型在 HFMD 的发病率的预测中均能较好地拟合, 对于发病的波峰和波谷有较准确的预测。两者相比较, SARIMA 模型具有更高的预测精度, 与其他基于 SARIMA 模型对区域内的 HFMD 发病率研究结果一致^[12-13]。

本研究发现, SARIMA (1, 0, 1) (0, 1, 1)₁₂ 模型参数与国内其他研究结论相同, 但预测的 MAPE 为 159.949%, 与其他研究相比误差值偏高^[14-15], Holt-Winters 加法模型的 MAPE 值同样偏高, 为 327.069%。误差增大的原因可能与训练集中 2020—2022 年实际 HFMD 发病率高存在下降趋势使得预测值偏低有关, 而测试集 2023 年 1—12 月的实际发病率中存在一个显著上升的高峰, 使得预测值与实际值两者相比的误差增大。

HFMD 的实际发病率与个体年龄、疫苗接种、经济条件、个人免疫力以及气候变化等多种因素均有关系^[16], 本研究的不足之处在于模型只纳入了时间这一个影响因素, 而非混合统计模型, 预测结果具有一定局限性。

综上所述, HFMD 发病具有季节性波动趋势, 构建的 SARIMA 模型在 HFMD 发病率预测中表现更优, 能更有效地拟合发病率序列并进行短期预测, 为 HFMD 的动态防治工作提供科学依据。

利益冲突 所有作者声明无利益冲突

参 考 文 献

- [1] ESPOSITO S, PRINCIPI N. Hand, foot and mouth disease: Current knowledge on clinical manifestations, epidemiology, aetiology and prevention[J]. Eur J Clin Microbiol Infect Dis, 2018, 37(3): 391-398.
- [2] 刘艳艳, 杨善蕊, 段忠文, 等. 2021 年云南省手足口病病例中柯萨奇病毒 A2 型的测序定型和基因特征分析[J]. 中华微生物学和免疫学杂志, 2023, 43(1): 27-34.
- [3] QI H C, LI Y, ZHANG J, et al. Quantifying the risk of hand, foot, and mouth disease (HFMD) attributable to meteorological factors in East China: A time series modelling study[J]. Sci Total Environ, 2020, 728: 138548.
- [4] YANG F, MA Y, LIU F, et al. Short-term effects of rainfall on childhood

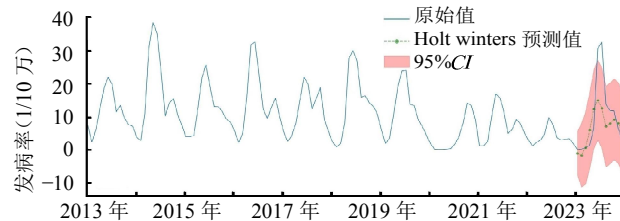


图 5 SARIMA(1, 0, 1)(0, 1, 1)₁₂ 模型预测图

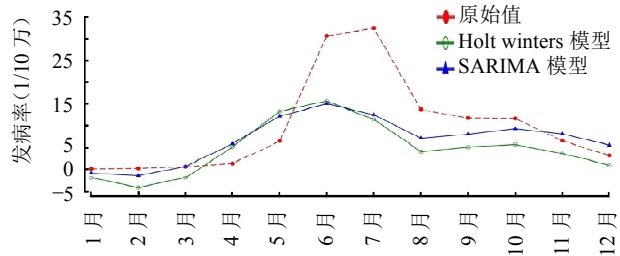


图 6 2023 年 1 ~ 12 月两种模型预测值与实际值对比

- hand, foot and mouth disease and related spatial heterogeneity: Evidence from 143 cities in mainland China[J]. BMC Public Health, 2020, 20(1): 1528.
- [5] 王萍萍. 人口总量略有下降, 城镇化水平继续提高[EB/OL]. 2023-01-18/2024-03-20 http://www.ce.cn/xwzx/gnsz/gdxw/202301/18/t20230118_38353400.shtml.
 - [6] 国家统计局. 中国统计年鉴 2023[M]. 北京: 中国统计出版社, 2023: 10.
 - [7] 徐梦蝶, 李海朋, 尹莉, 等. 基于 Holt-Winters 模型探讨新型冠状病毒肺炎疫情对连云港市流感样病例时间序列的影响[J]. 现代预防医学, 2022, 49(8): 1370-1374, 1395.
 - [8] 汤梦莹, 宋晓坤, 梁凯琼, 等. 三种预测模型对中国流行性腮腺炎月度发病率预测的应用与比较[J]. 实用预防医学, 2023, 30(11): 1392-1396.
 - [9] 陈春艳, 陈亿雄, 赵执扬, 等. SARIMA 模型和 LSTM 神经网络在预测深圳市宝安区手足口病疫情中的应用[J]. 山西医科大学学报, 2022, 53(10): 1302-1307.
 - [10] YU G C, FENG H F, FENG S, et al. Forecasting hand-foot-and-mouth disease cases using wavelet-based SARIMA-NNAR hybrid model[J]. PLoS One, 2021, 16(2): e0246673.
 - [11] TIAN C W, WANG H, LUO X M. Time-series modelling and forecasting of hand, foot and mouth disease cases in China from 2008 to 2018[J]. Epidemiol Infect, 2019, 147: e82.
 - [12] 袁永旭, 宋立群, 吴文婧, 等. 基于 ARIMA 模型的山西省手足口病月发病率的预测研究[J]. 中华医学图书情报杂志, 2022, 31(4): 20-26.
 - [13] 贺箫楠, 宋晓辉, 朱鑫. 基于 ARIMA 模型的洛阳市手足口病发病率预测[J]. 现代预防医学, 2019, 46(3): 403-406.
 - [14] YU C C, XU C J, LI Y H, et al. Time series analysis and forecasting of the hand-foot-mouth disease morbidity in China using an advanced exponential smoothing state space TBATS model[J]. Infect Drug Resist, 2021, 14: 2809-2821.
 - [15] 刘天, 张丽杰, 翁熹君, 等. SARIMA 模型参数设置探讨[J]. 实用预防医学, 2019, 26(12): 1530-1534.
 - [16] ZHANG R, LIN Z H, GUO Z, et al. Daily mean temperature and HFMD: Risk assessment and attributable fraction identification in Ningbo China[J]. J Expo Sci Environ Epidemiol, 2021, 31(4): 664-671.

收稿日期: 2024-03-26

(本文编辑: 钟美春)