

## · 调查报告 ·

## 时间序列分析法预测高温干旱期间住院病例数量

段 芹<sup>1</sup>, 余兰英<sup>1</sup>, 徐小利<sup>2</sup>, 鲁建央<sup>1</sup>, 刘达伟<sup>1</sup>, 钟朝晖<sup>1△</sup>

(1. 重庆医科大学公共卫生学院 400016; 2. 重庆市第五人民医院 400062)

**摘要:**目的 探讨重庆市局部地区高温干旱期间住院病例数量变化趋势,为今后建立与完善高温干旱期间住院病例数量预警系统提供依据。方法 通过分析重庆市局部地区高温干旱期间住院病例数量的变化规律,使用时间序列分析法中的 ARIMA 模型分析和预测住院病例的变化趋势。结果 建立了 ARIMA(p,d,q)模型,绝大多数实际值均落在两条预测值的 95% 置信线内,且该模型进行了短期的预测。结论 时间序列分析法用于高温干旱期间住院病例数量变化的研究中具有较好的适应性和适用性,ARIMA(p,d,q)模型能较好的反应并预测重庆市局部地区该期间病例数的变化趋势。

**关键词:**住院;高温干旱;时间序列;ARIMA 模型;趋势预测

doi:10.3969/j.issn.1671-8348.2011.30.028

文献标识码:A

文章编号:1671-8348(2011)30-3084-03

## Time series analysis for prediction of the inpatient-numbers during hot and drought months

Duan Qin<sup>1</sup>, Yu Lanying<sup>1</sup>, Xu Xiaoyang<sup>2</sup>, Lu Jianyang<sup>1</sup>, Liu Dawei<sup>1</sup>, Zhong Chaohui<sup>1△</sup>

(1. Department of Epidemiology, Chongqing Medical University, Chongqing 400016, China; 2. the Fifth People's Hospital of Chongqing, Chongqing 400062, China)

**Abstract:** Objective To investigate the variation trend of the number of inpatients during hot weather and drought period, in order to explore how to build and improve early warning system of the amount of inpatients during hot weather and drought months. Methods Through analyzing the variation regularity of time series, ARIMA(p,d,q) model was used to analyze the dynamic variation of the amount of inpatients. Results The model ARIMA(p,d,q) was established, the vast majority of the actual values fell to the 95% confidence limits of two predictive line, and it produced a good fitness to predict the amount of inpatients in a short future. Conclusion ARIMA(p,d,q) model was applicable and practical in analyzing the variation regularity of the amount of inpatient, which indicate the model ARIMA(p,d,q) can response and predict well the dynamic variation of the amount of inpatients.

**Key words:** hospitalization; hot weather and drought; time series; ARIMA model; tendency predication

近年来,夏季高温旱灾天气在全球的频繁发生,对人群的健康造成了极大的影响<sup>[1-4]</sup>,如何早期预警、有效促进人群健康已成为公共卫生领域亟待解决的重大问题。在中国重庆地区,7、8 月份副热带高压非常活跃,造成该地区夏季发生高温伏旱的概率较大。本次研究应用求和自回归移动平均模型(Autoregressive integrated moving average model, ARIMA)对重庆市局部地区高温干旱期间住院病例资料进行时间序列分析并建立预测模型,为今后建立与完善高温干旱期间人群住院病例数量预警系统提供依据,以便合理分配卫生资源,降低高温干旱气象条件对人群健康的不利影响。

**1 资料与方法**

**1.1 研究对象** 本次研究为回顾性研究,采用分层整群抽样的方法,收集 2006 年 7~8 月共 62 d 高温干旱期间重庆市巴南区与云阳县两所医院的全部住院病例资料。

**1.2 统计学处理** 采用 Microsoft Excel 2003 软件建立病例资料数据库,运用 SAS 8.2 统计软件中的统计学模块进行数据的处理与分析。

**1.3 原理**

**1.3.1 ARIMA 模型** ARIMA 模型即求和自回归滑动平均模型,该模型简记为 ARIMA(p,d,q),其中,p、d、q 分别表示自回归阶数、差分阶数、移动平均阶数。

**1.3.2 时间序列数据建模过程**可归纳为 4 个基本步骤 (1) 时序数据的预处理(平稳性检验和白噪声检验);(2)模型的识

别;(3)模型参数估计与检验;(4)模型的优化。建立 ARIMA 模型的前提条件是:作为预测对象的时间序列是一零均值的平稳随机序列。因此在建立模型前,需要对时间序列进行零均值化和差分平稳化处理。通过计算差分序列的自相关系数和偏相关系数,做 ACF 和 PACF 图并确定模型,检验 SACF 和 SPACF,分出模型的类型。用最小二乘法进行参数估计,通过统计量 QLB 对模型的检验,确定模型是否适当。并根据 AIC 和 SBC 信息准则来进行模型优化,即 AIC 函数或 SBC 函数达到最小值的模型被认为是相对最优模型,由此得出序列的最佳 ARIMA 模型,并应用该模型做出预测。最后的模型是一个多次反复验证、估计、检验并与充足的传统模式相统一的结果<sup>[5]</sup>。

**2 结果**

**2.1 住院患者数量时间数据的预处理** 按每日的住院病例数量依次统计,得到一组时间序列数据,其时序图见图 1,自相关函数图(ACF),偏自相关函数图(PACF)。对该时序进行平稳性检验和白噪声检验。平稳性检验中,图 1 可见序列值波动范围比较大,自相关系数图和偏自相关系数图均表现出一种拖尾性,没有很快衰减为零,所以该时间序列为一个非平稳序列;但对其作一阶差分后该序列已平稳。白噪声检验结果,可知当延迟时期为 6、12 h,检验统计量 QLB 分别为 20.31、41.12, P 值分别为 0.002 4 和 <0.000 01,即延迟期数小于等于 6 或 12 时的序列值之间至少存在自相关系数不为零情况,所以该序列为非白噪声序列。差分自相关图见图 2,差分偏自相关图见图 3。

**2.2 模型识别** 观察差分后的自相关系数图见图 2, 偏自相关系数图见图 3, 发现自相关系数 1 阶截尾, 偏自相关系数拖尾, 根据 ARMA 模型的自相关系数与偏自相关系数的性质, 初步判断出差分后序列为 MA(1) 模型。

**2.3 模型参数估计与检验** 条件最小二乘法估计参数示参数 MA1,1 显著非零 ( $P < 0.0001$ ), 参数 MA1,1 通过检验; 模型的统计学意义统计检验即残差序列的白噪声检验, 用统计量 QLB 作为模型检验的统计量延迟时期为 6 时, 检验统计量 QLB 为 2.54,  $P$  值为 0.7703, 即延迟期数小于等于 6 时残差序列值之间相互独立, 该残差序列为白噪声序列, 说明差分后的 MA(1) 模型即 ARIMA(0,1,1) 模型提取信息充分。

其 ARIMA(0,1,1) 模型的表达式为:  $(1-B)X_t = (1-\theta_1B)\epsilon_t$ , 由此得出该时间序列的模型为  $Y_t - Y_{t-1} = \epsilon_t - 0.86227 * \epsilon_{t-1}$ 。

对应的  $AIC=515.0224, SBC=519.2441$ 。

表 1 AR(6)模型的条件最小二乘估计

参数	系数的估计	标准差	t	P
MU	-0.219 62	0.484 54	-0.45	0.652 2
AR1,1	-0.909 31	0.127 52	-7.13	<0.000 1
AR1,2	-0.790 63	0.171 41	-4.61	<0.000 1
AR1,3	-0.532 15	0.193 29	-2.75	0.008 0
AR1,4	-0.448 49	0.193 72	-2.32	0.024 4
AR1,5	-0.341 11	0.172 43	-1.98	0.053 0

表 2 AR(2)模型的条件最小二乘估计

参数	系数的估计	标准差	t	P
MU	-0.169 91	1.021 15	-0.17	0.868 4
AR1,1	-0.756 52	0.118 65	-6.38	<0.001
AR1,2	-0.438 20	0.118 77	-3.69	0.000 5

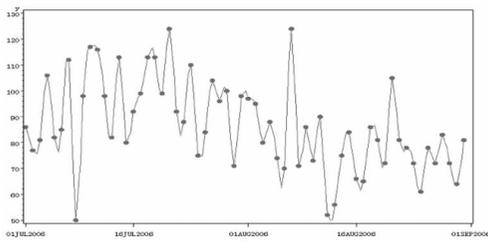


图 1 2006 年 7~8 月高温干旱期间住院病例数时序图

**2.4 模型的优化** 根据 AIC 和 SBC 准则, 观察图 5 尝试拟合 AR(6) 模型, 结果见表 1: 经  $t$  参数 AR1,5 的系数  $P > 0.05$ , 不显著; 去掉该系数重新拟合发现结果残差序列未通过白噪声检验, 所以, 该模型不适合本次研究序列。同理拟合其他模型, AR(2) 模型的拟合结果见表 2, 可知除均数外, 其余参数均显著非零 ( $P < 0.05$ ), 且残差序列通过了白噪声检验 ( $P > 0.05$ ), 说明该模型提取信息充分, 拟合成功。经计算, AR(2) 模型即 ARIMA(2,1,0) 模型对应的  $AIC=524.0788, SBC=530.4114$ 。比较上述拟合成功的 ARIMA(2,1,0) 模型和 ARIMA(0,1,1) 模型的优劣见表 3, 可知模型 ARIMA(0,1,1) 拟合本次研究的序列较好。故最后的模型表达式为:  $Y_t - Y_{t-1} = \epsilon_t - 0.$

$86227 * \epsilon_{t-1}$ 。

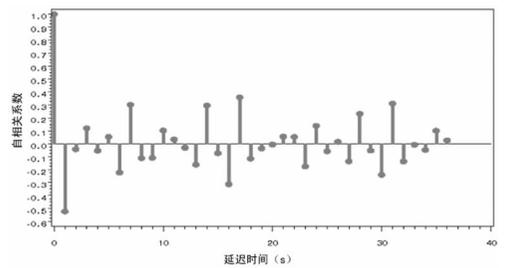


图 2 2006 年 7~8 月高温干旱期间住院病例数一阶差分自相关系数图

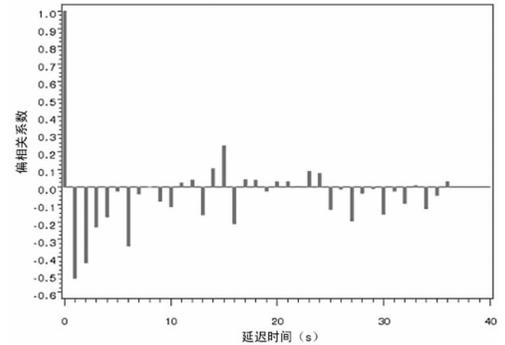


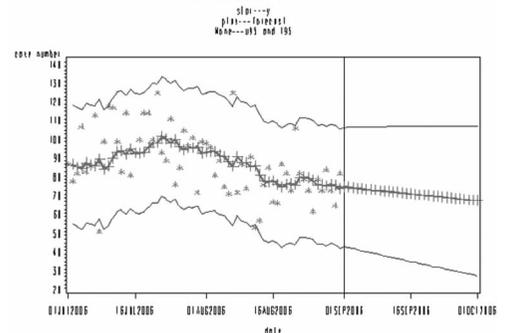
图 3 2006 年 7~8 月高温期间住院病例数一阶差分后偏自相关系数图

表 3 两种拟合模型优劣比较

模型	AIC	SBC
ARIMA(2,1,0)	524.0788	530.4114
ARIMA(0,1,1)	515.0224	519.244

**2.5 模型 ARIMA(0,1,1) 预测效果评价** 用 ARIMA(0,1,1) 模型对 2006 年 7~8 月每日的住院病例数量进行模拟并同时预测 2006 年 9 月每日的住院病例数量见图 4, 绝大多数实际值均落在两条预测值的 95% 置信线内, 可知该模型与实际值拟合较好, 同时, 根据时间序列的延续性, 该模型在一定范围内可对高温期间住院病例数进行预测。

Forecast Chart



星号示实际值; 加号示预测值; 实线示预测值的上下 95% 置信线。

图 4 2006 年 7~8 月住院病例数的实际值、预测值、上下 95% 置信线的时间序列图

3 讨论

全球气候变暖、大气环流异常、过度城市化, 使得全球的极端高温天气频繁发生, 并呈现强度大、频次高、范围广等特点,

且极端的高温天气常伴随着严重的干旱,例如 2006 年重庆、四川出现百年一遇的高温干旱天气。据有关部门统计,7 月中旬至 8 月下旬,重庆市大于或等于 38℃ 的高温日数达 21 d,创历史新高,中暑人数达万余人;四川全省受灾人数超过 2 328.31 万。持续异常的高温干旱天气对人们的生理适应能力带来严峻挑战,从而对人群健康产生直接或间接的影响,其直接影响主要使人群患某些疾病,如某些虫媒传染病,或使已患的疾病加重,如心脑血管疾病等<sup>[6]</sup>,从而造成死亡率大大增加<sup>[7]</sup>;间接影响主要对某些与人群健康有关的社会因素产生影响,如造成经济损失,加重环境污染等。

面对这种突发公共卫生事件,人类能够做到的最有效的办法,就是在灾害降临之前构筑一道坚固的公共卫生防御屏障,建立和健全快速应急响应体系,预防和减少疾病的发生和流行<sup>[8]</sup>。因此,监测与预警关系到突发事件应急处理的成败。在重庆,由于各种内外因素的影响,夏季发生高温干旱天气的概率较大,在该市建立起高温干旱的疾病的预警系统具有重大意义。

时间序列法是一种考虑变量随时间发展变化规律并用该变量以往的统计资料建立数学模型作外推的预测方法。该方法近年来已被广泛应用于公共卫生研究领域,如国内的乙型肝炎的季节性分布特征、流行性出血热和血吸虫病的发病预测等<sup>[9-11]</sup>,国外的新加坡的三级医院在 SARS 期间病床数目预测和监测、巴基斯坦卡拉奇男性献血者志愿者丙型肝炎病毒血清阳性短期预测、医院急诊部门的人流量预测<sup>[12-15]</sup>。

本研究用该预测法中重要且预测精度较高的 ARIMA 模型对 2006 年高温干旱期间 7~8 月共 62 d 重庆市两所医院的住院病例资料进行分析,以短期预测住院人数变化情况。结果表明,高温干旱期间的住院病例数量所形成的时间序列为非白噪声序列,即每日住院患者数量之间有一定的联系性,这可用疾病中以慢性疾病为主、某些疾病的传染性及类似的生活环境和气象因素对疾病的影响等来解释;另外,该期间的住院病例虽为非平稳序列,但经一阶差分后其序列值平稳。说明该方法用于高温干旱期间住院病例数量变化的研究中具有较好的适应性,为此期间预警系统的建立与完善提供了新思路。此外,通过模型识别、模型及参数检验、模型的模拟和优化等,最终找到了能够很好反应该期间住院病例数量变化情况及其变化趋势的模型,即 ARIMA(0,1,1)模型。

准确的分析和较为准确预测高温期间住院人数的动态趋势,可为更合理地分配卫生资源提供参考依据,从而降低高温干旱气象条件对人群健康的不利影响。因此,运用 ARIMA 模型预测高温期间住院人数具有重要的实际意义。该模型对样本容量和概率分布没有严格要求,但是,在运用时仍要注意以下几个问题<sup>[15]</sup>:(1)在模型识别时候需要 50 个以上历史统计数据,这对按月、季或年记录的资料往往较难收集,因此,在应用此方法时,收集数据资料是一项十分费时费力的工作。(2)在建模时,逐次增加模型的阶数,的确有可能达到使模型通过统计检验的要求。但是提高模型的阶数,就要增加模型中的项数,对变量前的系数在实际意义上往往解释不通,或者根本就无从分析,因此重点应用于 1 阶、2 阶模型。(3)该方法进行预测所依据的基本假设是一个时间序列的未来发展模式与过去的模式是一致的。对未来的短期预测,这一个假设往往是可以

满足的,但是对未来长期的预测,这一假设显然是不符合实际的。通常用来预测的时间段不能超过历史资料时间段的 1/4。

本次研究选择了 62 个时间点的资料,符合 ARIMA 模型对资料时间点数量上的要求,但是,ARIMA 模型只适用于短期预测,本次研究分析的资料是 2006 年重庆高温干旱比较严重期间的住院资料,因此,利用该资料只能短期预测高温期间的住院患者数量,而不能无限推广到非高温期的预测。

#### 参考文献:

- [1] 李永红,陈晓东,林萍.高温对南京市某城区人口死亡的影响[J].环境与健康杂志,2005,22(1):6-8.
- [2] Kovats RS, Hajat S, Wilkinson P, et al. Contrasting patterns of mortality and hospital admissions during hot weather and heat waves in Greater London, UK[J]. Occup Environ Med, 2004, 61(11):893-898.
- [3] 周汉勇,杜晓刚,王大尧,等.重症中暑并发多器官功能障碍综合征 1 例[J].重庆医学,2006,35(24):2264-2264.
- [4] Kovats RS, Kristie LE. Heatwaves and public health in Europe[J]. Eur J Public Health, 2006, 16(6):592-599.
- [5] Ljung GM, Box GEP. On a measure of lack of fit in time-series models[J]. Biometrika, 1978, 65:297-303.
- [6] 刘学恩,李群娜,赵宗群.气温及冷空气对武汉市心脑血管疾病死亡率的影响[J].中国公共卫生,2002,18(8):948-950.
- [7] 李永红,陈晓东,林萍.高温对南京市某城区人口死亡的影响[J].环境与健康杂志,2005,22(1):6-8.
- [8] 李璐.重庆市突发公共卫生事件应急系统研究[J].重庆医学,2007,36(18):1903-1904.
- [9] 李来英,张小平.起伏型时间序列分析方法在流行性出血热预测中的应用[J].中国卫生统计,1997,14(3):64.
- [10] 赛晓勇,张治英,徐德忠,等.不同时间序列分析法在洞庭湖区血吸虫病发病预测中的比较[J].中华流行病学杂志,2004,25(10):863-866.
- [11] 吴家兵,叶临湘,尤尔科,等.时间序列模型在传染病发病率预测中的应用[J].中国卫生统计,2006,23(3):276.
- [12] Earnest A, Chen MI, Ng D, et al. Using autoregressive integrated moving average (ARIMA) models to predict and monitor the number of beds occupied during a SARS outbreak in a tertiary hospital in Singapore[J]. BMC Health Serv Res, 2005, 5:36.
- [13] Akhtar S, Rozi S. An autoregressive integrated moving average model for short-term prediction of hepatitis C virus seropositivity among male volunteer blood donors in Karachi, Pakistan[J]. World J Gastroenterol, 2009, 15(13):1607-1612.
- [14] Sun Y, Heng BH, Seow YT, et al. Forecasting daily attendances at an emergency department to aid resource planning [J]. BMC Emerg Med, 2009, 9:1.
- [15] 徐国详.统计预测和决策[M].上海:上海财经大学出版社,2004,174-176.