

两种模型预测北京市痢疾发病率的效果比较

田祎,李锡太,贾蕾,王小莉,钱海坤,吴晓娜,李爽,王全意*

摘要:目的 比较 Winters 乘法预测模型和自回归移动平均模型(Autoregressive Integrated Moving Average Model, ARIMA)两种模型对北京市痢疾发病率的预测效果,从而选择合适的预测方法。方法 收集2007~2012年北京市痢疾分月发病率资料,分别运用 Winters 乘法预测模型和 ARIMA 模型进行建模,评估两种模型的预测效果,并用较好的模型预测2013年第一季度北京市痢疾发病率。结果 以2012年数据验证两种模型的拟合情况,预测误差百分比比较, Winters 乘法(1.13%)优于 ARIMA 模型(6.80%)。Winters 乘法预测2013年1、2、3月痢疾发病率分别为1.82/10万、1.54/10万、1.85/10万。结论 Winters 乘法预测模型可以较好地反映北京市痢疾发病趋势,用于预测预警。

关键词: Winters 乘法预测模型; ARIMA 模型; 痢疾

中图分类号: R531.3 文献标识码: A 文章编号: 1009-9727(2013)7-841-03

Comparing the results of two models in prediction of dysentery incidence. TIAN Wei, LI Xi-tai, JIA Lei, et al. (Beijing Municipal Centers for Disease Control and Prevention, Beijing 100013, P. R. China)

Abstract: Objective To compare the differences between winters multiplication model and autoregressive integrated moving average model(ARIMA) in predicting the incidence of dysentery in Beijing. Methods The monthly incidence data of dysentery from January 2007 to December 2012 in Beijing were collected and modeling the data with winters multiplication model and ARIMA. The results of predicting the incidence of dysentery in the first quarter of 2013 in Beijing were evaluated. Results After the assessment of fit of these two models using data in 2012, measured by prediction percentage error, winters multiplication model(1.13%) was found to be better than ARIMA (6.80%). The predicting incidence rates of dysentery by using the winters multiplication model in the first quarter of 2013 were 1.82/100000, 1.54/100000 and 1.85/100000. Conclusions winters multiplication model could well reflect the trend of the incidence of dysentery in Beijing and it was suitable for predicting the future trend dysentery.

Key words: Winters multiplication prediction Model; Autoregressive Integrated Moving Average Model; Dysentery

痢疾是指以腹部疼痛、里急后重、脓血便为主要症状的肠道传染性疾病,严重时可伴有全身中毒等症状,属乙类传染病,多发于夏秋季节,冬春两季也可见。2004年1月1日起,我国启用“中国疾病预防控制中心信息管理系统”,将痢疾(包括细菌性痢疾和阿米巴痢疾)纳入传染病监测内容,监测发病和死亡情况,数据来自于各哨点医院。不同的数据需要不同的预测模型和分析方法,使预测值与实测值拟合较好,提高预测精准度。此次研究的目的是比较两种不同模型预测效果,找出较好的预测北京市痢疾发病率方法,对北京市痢疾发病情况进行及时准确地预测预警。

1 资料与方法

1.1 资料来源 2007年1月至2012年12月北京市痢疾发病率资料来源于《疾病监测信息报告管理系统》。

1.2 方法和原理 (1)Winters 乘法预测模型,亦称“季节指数预测法”。原理为先建立一个能描述整个时间序列总体发展趋势的数学方程,然后考虑季节变化对预测对象的影响,计算出季节指数,最后再将两者结合,从而得到能够描述总体发展趋势与季节性变动的

预测模型,用于预测^[1]。

平滑方程可以表示为 $S_t = a(Y_t/F_{t-L}) + (1-a)(S_{t-1} + b_{t-1})$; $b_t = b(S_t - S_{t-1}) + (1-b)b_{t-1}$; $F_t = g(Y_t/S_t) + (1-g)F_{t-L}$ ^[2]。其中 S_t 为稳定性项; b_t 为趋势项; F_t 代表季节性; L 为季节性(或周期项),如月数; Y_t 为实测值; X_t 为 t 时刻的实测值; a 、 b 、 g 为平滑系数,取值范围为(0,1),一般根据经验取值,当实际值与预测值的相对误差和最小时,则认为此时平滑系数组合最优。

(2)ARIMA 模型是以数理统计和微积分等传统数学理论为基础建立起来的预测模型,属于时间序列分析方法。利用事物发展具有一定惯性(即延续性)的原理,根据预测变量自身在各时刻的变化规律,以时间 t 综合替代各种影响因素,建立时序模型以达到预测未来的目的。ARIMA 模型的形式可表示为 $ARIMA(p, d, q) \times (P, D, Q)$, 其中 p 、 d 、 q 分别表示自回归阶数、差分阶数和移动平均阶数, P 、 D 、 Q 分别为季节性自回归阶数、季节性差分次数和移动平均阶数, s 为季节周期。模型构建过程包括数据时间序列平稳化、模型识别、参数估计和模型诊断四个步骤^[3]。

作者单位:北京市疾病预防控制中心,北京 100013

作者简介:田祎(1987~),女,河北沧州人,硕士研究生,公共卫生医师,主要从事疾病控制研究。

*通讯作者: E-mail: @bjcdcxm@126.com

1.3 数据分析方法 采用SPSS18.0软件包建立数据库、建模,并对数据进行分析。a值选择0.05为是否具有统计学差异的临界值。

2 结果

2.1 北京市痢疾2007~2012年报告情况 痢疾发病存在明显的季节性,其中每年的7、8月份发病率均为全年的最高点。6年中,2007年8月发病率最高,为55.67/10万,最低为2012年2月(2.10/10万)。同月份发病率与上一年比较,每年均有不同程度下降。

2.2 两种模型预测结果

2.2.1 Winters乘法模型预测 SPSS预测模块中专家建模,自动选择了Winters乘法模型,结果显示,平稳的R²为0.813, Ljung-Box Q检验统计量为26.82, P=0.030, BIC=1.512,模型拟合数据较好,见图1。

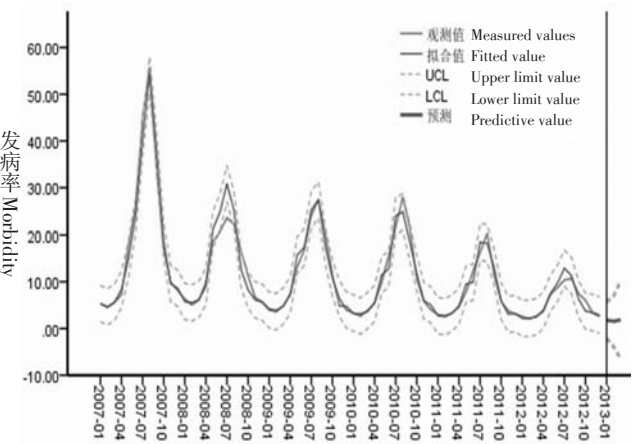


图1 Winters乘法预测趋势图

Figure 1 The predicting trend chart by winters multiplication model

2.2.2 ARIMA模型预测 利用常用于痢疾预测的

ARIMA模型建模,结果显示,模型选择(0,0,0)×(0,1,0)时,模型最优。平稳的R²=0.564, R²=0.564,残差序列的Ljung-Box Q检验统计量为75.49, P=0.000,即在所有点上均无统计学差异,据此认为残差序列为白噪声。正态化的BIC=4.419,常数项估计值为8.692。见图2。

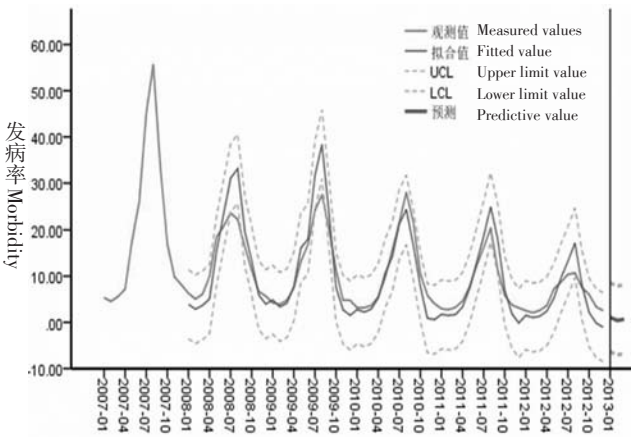


图2 ARIMA乘法预测趋势图

Figure 2 The predicting trend chart by autoregressive integrated moving average model

2.2.3 两种模型预测比较 以2012年数据为例,验证两种模型的拟合情况,见表1。Winters乘法预测模型预测误差百分比为1.13%,低于ARIMA模型,拟合数据较好。

2.4 Winters乘法预测模型预测结果 运用Winters乘法预测模型预测2013年第一季度痢疾发病率,结果整理见表2。结果显示2013年第一季度痢疾发病率均较低,其中2月份低于其它两个月。

表1 两种模型发病率拟合情况比较
Table1 Comparison between two models for incidence fitting

月份 Month	实测值 Measured values	预测值(1/10万) Predictive value		预测误差百分比(%) Prediction error percentage	
		Winters乘法 Winters multiplication Model	ARIMA模型 ARIMA Model	Winters乘法 Winters multiplication Model	ARIMA模型 ARIMA Model
1月	2.54	2.17	1.54	0.38	1.00
2月	2.09	2.19	1.05	-0.11	1.04
3月	2.63	2.51	1.34	0.11	1.29
4月	3.65	3.92	2.44	-0.27	1.21
5月	7.27	7.51	5.31	-0.24	1.96
6月	8.74	9.67	9.60	-0.93	-0.86
7月	10.39	12.90	13.23	-2.51	-2.85
8月	10.70	11.32	17.19	-0.62	-6.49
9月	7.46	6.36	7.99	1.10	-0.53
10月	6.07	3.74	2.15	2.33	3.92
11月	3.35	3.41	-1.13	-0.06	3.48
12月	2.59	2.90	-1.04	-0.31	3.63
合计 Total	67.48	68.60	60.67	-1.13	6.80

表2 Winters 乘法模型预测北京市2013年第一季度痢疾发病率(1/10万)

Table 2 Using Winters multiplication model to predict the incidence of diarrhea in the first quarter of 2013(1/100,000)

月份 Month	Winters 乘法 Winters multiplication Model		
	下限值	预测值	上限值
	Upper limit value	Predictive value	Lower limit value
1月 January	0	1.82	5.70
2月 February	0	1.54	6.96
3月 March	0	1.85	10.20

3 讨论

痢疾发病率的监测,可以有助于在早期预测到疾病的发病趋势,对传染病的流行、暴发进行预警,及时采取防控措施,防止痢疾疫情的流行蔓延,不但能减轻疾病带来的痛苦,也能节约成本。痢疾发病率与多方面因素有关,有数据表明,北京市痢疾每年发病率均居于全国前5位,但建国以来,北京市痢疾发病率随着社会、经济、卫生事业等方面的发展出现了明显下降^[4]。2007~2012年北京市痢疾发病情况显示:痢疾发病存在明显的季节性,5~10月为高发月份,病例数占全年的80%左右,7、8月份为全年高峰;发病率呈逐年稳定下降趋势,同月份发病率与上一年比较,每年均有不同程度下降。

传染病预测可以提前识别疾病高峰来临的迹象,有助于做好充分的物质和技术储备,防止暴发事件的发生。时至今日,传染病预测方法已多达几十种^[5-6],如ARIMA模型、GM(1,1)模型、时间序列模型^[7]、Serfling回归模型、BP网络预测等。由于任何预测模型都有它的适用范围和优缺点,故还没有一种模型,既适用于任何传染病的预测模型,又有极高的预测精度。因此,在实际工作中应就某一特定问题进行深入研究,探讨最合适的预测方法。ARIMA模型和Winters乘法模型均可以用于呈周期性变化的数据,进行短期预测。此次研究用两种模型分别预测2013年第一季度痢疾的发病率,并以2012年数据为例,验证两种模型的拟合情况可以发现,Winters乘法模型预测误差百分低于ARIMA模型,预测值与实测值拟合更好,这可能因为Winters乘法模型可以将时间趋势与季节因素结合,考虑时间序列的季节波动^[8],而ARIMA模型更适合用于处理有趋势且季节平稳的数据,故Winters乘法模型预测更适用于北京地区痢疾发病率的预测。预测发现2013年1~3月痢疾发病率较低,且低于往年平均水平,其中2月份最低。

实际工作中,传染病预测可以及早发现传染病发展趋势,为预警奠定基础,亦为制定防治措施及策略

提供理论依据,而预警是预测的实际运用。以此次研究为例,我们可以及时发现可能暴发,如图1所示,当在发病率预测趋势图中监测到,观测值(图1中红线实线表示)高于拟合值(蓝线实线表示)时,标志异常信号点出现,提示痢疾暴发或者发病高峰即将来临,即发出预警信号。根据预警信号,可以建议相关卫生行政部门做好各部门的协调工作,实现信息共享和利用、疫情共同处置等,对切实做好北京市痢疾防控工作有重要指导意义。

参考文献:

- [1] Mao Y, Jiang L, Yin ZW. Research of Winters Prediction Model Based on Parallel Computing[J]. Sci Technol Eng, 2012, 12(20): 5023-5026. (In Chinese)
(茅颖, 姜麟, 尹正文. 温特斯模型参数并行求解方法与应用[J]. 科学技术与工程, 2012, 12(20): 5023-5026.)
- [2] Zhang Y, Yang LY. On the Applications of the Additive Model and Multiplicative Model of Time Series Analysis[J]. Statistics & Information Tribune, 2005, 20(4): 45-47. (In Chinese)
(张颖, 杨兰英. 时间数列分析中的加法模型与乘法模型[J]. 统计与信息论坛, 2005, 20(4): 45-47.)
- [3] Lin J, Wu HY, Li YT, et al. Establishing and applying of autoregressive integrated moving average model to predict the incidence rate of dysentery in Shanghai[J]. Chin J Prev Med, 2010, 44 (1): 48-53. (In Chinese)
(黎健, 吴寰宇, 李燕婷, 等. 上海市痢疾发病率预测自回归求和移动平均模型的构建与应用[J]. 中华预防医学杂志, 2010, 44(1): 48-53.)
- [4] Wu XN, Jia L, Liu GR, et al. Analysis of the epidemiological characteristics of dysentery-like cases from 2004 to 2009 in Beijing[J]. Chin J Dis Control Prev, 2011, 15(10): 872-875. (In Chinese)
(吴晓娜, 贾蕾, 刘桂荣, 等. 北京市2004~2009年痢疾样病例流行病学特征分析[J]. 中华流行病学杂志, 2011, 15(10): 872-875.)
- [5] Zhang Yi, Peng XM, Lu GL, et al. Application of three models to predict the influenza activity of 2011-2012 [J]. Internatl J Virol, 2011, 18 (6): 175-178. (In Chinese)
(张奕, 彭晓晏, 卢桂兰, 等. 应用三种模型对2011~2012流感病毒活动强度的预测研究[J]. 国际病毒学杂志, 2011, 18 (6): 175-178.)
- [6] Wang Xin, Cheng XW, Fang SS, et al. Application of exponential smoothing model in forecasting influenza incidence [J]. China Trop Med, 2011, 11(8): 938-939. (In Chinese)
(王昕, 程小雯, 房师松, 等. 指数平滑模型在流感样病例预测中的应用[J]. 中国热带医学, 2011, 11(8): 938-939.)
- [7] Yuan LY, Ding HL, Zhou YJ, et al. Prediction of bacillary dysentery incidence in Ankang City with time series model [J]. China Trop Med, 2011, 11(9): 1051-1053. (In Chinese)
(原凌云, 丁红玲, 周以军, 等. 应用时间序列模型预测安康市细菌性痢疾的发病率[J]. 中国热带医学, 2011, 11(9): 1051-1053.)
- [8] Sun Q, Yuan ZA, Tao FF, et al. Application of the prediction model of winters method-multiplicative in respiratory syndrome surveillance[J]. Chin J Dis Control Prev, 2011, 15(10): 905-908. (In Chinese)
(孙乔, 袁政安, 陶芳芳, 等. 温特斯乘法模型在呼吸道症候群监测中的应用[J]. 中华疾病控制杂志, 2011, 15(10): 905-908.)

收稿日期: 2013-01-22 编辑: 符式刚