

·论 著·

GM(1,1)季节指数模型在菌痢和阿米巴痢疾发病率预测中的应用

杨亮

山东医学高等专科学校,山东 临沂 276012

摘要: **目的** 探讨GM(1,1)季节指数模型在菌痢和阿米巴痢疾季度发病率预测中的应用价值。**方法** 基于2009—2013年我国菌痢和阿米巴痢疾各季度发病率资料建立GM(1,1)季节指数模型,用2014年各季度的数据验证模型,并与GM(1,1)模型预测效果比较,最后选择较优模型预测2015年的季度发病情况。**结果** GM(1,1)季节指数模型预测精度指标均方误差(MSE)、绝对平均误差(MAE)及相对平均误差(MAPE)均小于GM(1,1)模型,GM(1,1)季节指数模型预测2015年我国菌痢和阿米巴痢疾的季度发病率分别为1.4112/10万、3.0674/10万、4.5482/10万、2.0865/10万。**结论** GM(1,1)季节指数模型对菌痢和阿米巴痢疾季度发病率的预测效果优于GM(1,1)模型,预测应用价值较高。

关键词: GM(1,1)季节指数模型;GM(1,1)模型;菌痢;阿米巴痢疾

中图分类号: R516 **文献标识码:** A **文章编号:** 1009-9727(2015)08-951-04 DOI:10.13604/j.cnki.46-1064/r.2015.08.12

Application of GM (1, 1) season- exponential model in predicting incidence of bacillary dysentery and amebicdysentery

YANG Liang

Shandong Medical College, Linyi 276012, Shandong, P.R. China

Abstract: Objective To explore the application value of GM (1, 1) season- exponential model in predicting the incidence rate of bacillary dysentery and amebicdysentery per quarter in mainland China. **Methods** The GM(1, 1) season- exponential model was developed based on the quarterly incidence of bacillary dysentery and amebicdysentery in mainland China from 2009 to 2013, after the assessment of the model using quarterly data of 2014, the predicting effect was contrasted with GM(1, 1) model. The predicting results of the quarterly incidence rate of bacillary dysentery and amebicdysentery in 2015 were evaluated by better model. **Results** The mean square error (SE), mean absolute error (MAE) and mean absolute percentage error (MAPE) of predicting precision indexes by GM (1, 1) season- exponential model were all lower than those obtained from GM(1, 1) model. The predicted quarterly incidence rates of bacillary dysentery and amebicdysentery by GM (1, 1) season- exponential model in 2015 were 1.411 2/100 000, 3.067 4/100 000, 4.548 2/100 000, 2.086 5/100 000. **Conclusion** The application value of GM (1, 1) season- exponential model in predicting the incidence of bacillary dysentery and amebicdysentery were all better than the GM(1, 1) model.

Key words: GM(1, 1) season- exponential model; GM(1, 1) model; Bacillary dysentery ; Amebicdysentery

传染病预测是传染病防控过程中的关键环节,其意义在于为传染病防控部门制定应对策略提供科学依据^[1-2],因此传染病数理统计预测方法的研究成为目前疾病控制工作的关注热点。对于法定报告的传染病发病时间序列资料,灰色模型是常见的建模预测方法之一,其中GM(1,1)模型应用比较广泛,该模型是单变量一阶线性模型,由于对样本含量要求不高、不考虑概率分布规律,将原始数据累加为有规律的数据后建模预测,实用性较强。但传染病发病数据序列往往同时具备趋势性和季节性波动的特点,单一灰色模型仅能反映序列的总体趋势,往往无法准确刻画这种复杂、多变的趋势,导致预测精度不高^[3]。为此具体应用时只好增大采样间隔,以年为周期收集数据,其结果必然掩盖疾病的季节性,无法获取分月或季度的

预测结果,使预测应用价值大打折扣。GM(1,1)季节指数模型,应用季节指数对原始序列进行修正,实现GM(1,1)模型的优化,很好地解决了这个问题。

本文尝试将灰色GM(1,1)季节指数模型应用于菌痢和阿米巴痢疾发病趋势预测,并与灰色GM(1,1)模型的预测效果进行比较,探讨灰色GM(1,1)季节指数模型在菌痢和阿米巴痢疾预测中的应用价值。

1 资料与方法

1.1 资料来源 2009—2014年我国菌痢和阿米巴痢疾报告病例资料,来自国家卫计委网站法定报告传染病公告。人口资料源于2010—2014年国家统计局年鉴和国家统计局。

1.2 方法

1.2.1 研究方法 利用统计软件 SPSS20.0 和 Ex-

cel2003,对中国菌痢和阿米巴痢疾报告季发病率资料进行GM(1,1)季节指数模型拟合分析。本研究利用2009—2013年各季度报告发病率数据作为建模样本,用2014年1—4季度的数据验证模型,并与GM(1,1)模型预测效果比较,最后选择较优模型对2015年各季度发病率进行预测。

1.2.2 GM(1,1)季节指数模型构建^[3-10] 对于同时具有趋势性和季节波动的时间序列资料,利用季节指数对原始序列进行修正达到去除季节波动的目的,然后利用GM(1,1)模型进行分析预测,对修正序列的预测结果再进行季节指数还原便可得到原始序列的预测值。其具体构建步骤如下。

1.2.2.1 季节指数计算和原始序列修正 设原始序列 $x^{(0)}_i (i=1, 2, \dots, n)$,移动平均法剔除趋势值后计算季节指数 S_i 。按(1)式去除季节性波动得修正序列 $x^{(0)}_{i0}$

$$x^{(0)}_{i0} = x^{(0)}_i / S_i \quad (i=1, 2, \dots, n) \quad (1)$$

1.2.2.2 GM(1,1)模型建立 将修正序列按(2)式转换成1阶累加序列 $x^{(1)}_i$,强化其规律性。

$$x^{(1)}_i = \sum_{t=1}^i x^{(0)}_t \quad (i=1, 2, \dots, n) \quad (2)$$

对累加序列按(3)式计算紧邻均值序列 $z^{(1)}_i$ 。

$$z^{(1)}_i = 0.5[x^{(1)}_i + x^{(1)}_{i-1}] \quad (i=2, \dots, n) \quad (3)$$

$x^{(1)}_i$ 的灰微分方程表达式如下:

$$x^{(0)}_i + az^{(1)}_i = b \quad (4)$$

式中 a 、 b 分别称发展系数和灰作用量,为待定系数,(4)式的白化方程表达式为: $dx^{(1)}_i/dt + ax^{(1)}_i = b$ 。

定义灰参数向量 $\hat{\alpha} = [a, b]^T$,按(5)式以最小二乘法求得灰参数的解。

$$\hat{\alpha} = [a, b]^T = (B^T B)^{-1} B^T Y_n \quad (5)$$

$$\text{式中: } B = \begin{bmatrix} -z^{(1)}_{(2)} & 1 \\ -z^{(1)}_{(3)} & 1 \\ \vdots & \vdots \\ -z^{(1)}_{(n)} & 1 \end{bmatrix}, \quad Y_n = \begin{bmatrix} x^{(0)}_{(2)} \\ x^{(0)}_{(3)} \\ \vdots \\ x^{(0)}_{(n)} \end{bmatrix}$$

白化方程的解又称时间响应函数,式为: $\hat{x}^{(1)}_{i+1} = [x^{(0)}_{i+1} - b/a]e^{-at} + b/a \quad (6)$

据此求导还原模型得出修正序列拟合模型为:

$$\hat{x}^{(0)}_{i+1} = \hat{x}^{(1)}_{i+1} - \hat{x}^{(1)}_i = (1 - e^{-a})(x^{(0)}_{i+1} - b/a)e^{-at} \quad (i=1, 2, \dots, n-1) \quad (7)$$

将 $t(i \leq n-1)$ 代入(7)式求得修正序列拟合值,用后验差比值 C 、小误差概率 P 检验模型拟合效果。

设残差 $\varepsilon_i(t) = x^{(0)}_i - \hat{x}^{(0)}_i$, 则: $C = s_2/s_1$, $P = P\{|\varepsilon_i - \bar{\varepsilon}| < 0.6745s_1\} = \sum k/n$

式中:

$$s_1 = \sqrt{\frac{1}{n} \sum_{i=1}^n (x^{(0)}_i - \bar{x})^2}, \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x^{(0)}_i, \quad s_2 = \sqrt{\frac{1}{n} \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2},$$

$$\bar{\varepsilon} = \frac{1}{n} \sum_{i=1}^n \varepsilon_i, \quad \sum k \text{ 为 } |\varepsilon_i - \bar{\varepsilon}| < 0.6745s_1 \text{ 例数总和。}$$

根据 C 、 P 计算结果,参照表1综合判定模型拟合精度等级。如果拟合精度在勉强合格及以上等级时,可进行外推预测。

表1 灰色模型拟合精度检验等级参照表

Table 1 Reference table of precision test level of GM fitted values

等级 Rank	C	P
优秀 Excellency	$C \leq 0.35$	$P \geq 0.95$
合格 Qualification	$0.35 < C \leq 0.50$	$0.80 \leq P < 0.95$
勉强合格 Marginal qualification	$0.50 < C \leq 0.65$	$0.70 \leq P < 0.80$
不合格 Disqualification	$C > 0.65$	$P < 0.70$

1.2.2.3 预测 未来时间的拟合值称为预测值,故修正序列拟合值模型限定 t 的取值范围后既为预测值公式: $\hat{x}^{(0)}_{i+1} = \hat{x}^{(1)}_{i+1} - \hat{x}^{(1)}_i = (1 - e^{-a})(x^{(0)}_{i+1} - b/a)e^{-at} \quad (i=n, n+1, \dots, n+k-1) \quad (8)$

将修正序列预测值按对应季节指数还原便得原始序列预测值,表达式如下: $\hat{x}^{(0)}_i = \hat{x}^{(0)}_{i0} \times S_i \quad (9)$

1.2.2.4 预测精度评估 常用如下指标:均方误差

$$MSE = \sum_{i=1}^N [x^{(0)}(t) - \hat{x}^{(0)}(t)]^2 / N, \quad \text{绝对平均误差 MAE} =$$

$$\sum |x^{(0)}(t) - \hat{x}^{(0)}(t)| / N, \quad \text{相对平均误差 MAPE} =$$

$$\frac{1}{N} \sum_{i=1}^N \left| \frac{x^{(0)}(t) - \hat{x}^{(0)}(t)}{x^{(0)}(t)} \right|, \text{公式中 } N \text{ 为预测步数。}$$

2 结果

2.1 我国菌痢和阿米巴痢疾发病情况 2009—2014年我国共报告菌痢和阿米巴痢疾1 307 111例,各季度发病率存在波动,每年以第3季度的为最高,存在明显的季节性升高,发病率总体呈缓慢下降趋势。见图1。

2.2 季节指数计算及GM(1,1)建模 移动平均法计算季节指数不再赘述,直接应用SPSS20.0的季节分解模块得到各季度季节指数(0.4847、1.0768、1.6625、0.7760)和去除季节性波动的修正序列。

对修正序列进行累加和均值生成,用Excel2003按(5)式计算,得灰参数解为 $a=0.0228$, $b=5.3556$,并代入(6)式得修正序列时间响应函数为: $\hat{x}^{(1)}_{i+1} = -229.8323e^{-0.0228t} + 234.8947$,代入(7)式得修正序列拟合模型: $\hat{x}^{(0)}_{i+1} = 5.3004e^{-0.0228t} \quad (10)$,GM(1,1)模型计算结果见表2。

2.3 模型拟合效果检验 将时间序号 $t=1, 2, \dots, 19$ 代入(10)式, 求出拟合值(图2), 从图2可知2009—2013年各季度修正发病率值和拟合值基本吻合, 差别不大。拟合效果检验, 算得 $s_1=0.5876$, $s_2=0.2551$, 则 $C=0.4341$, $P=0.8421$, 据表1知模型拟合效果精度等级合格, 所建GM(1,1)模型可以直接外推预测。

2.4 预测及精度评估 将时间序号20、21、22、23分别代入(10)式计算修正序列预测值, 再按(9)式还原便得2014年菌痢和阿米巴痢疾各季度发病率的预测值依次为1.6284/10万、3.5359/10万、5.3360/10万、2.4347/10万。

对一个模型的预测精度评价, 仅考察MSE、MAE

和MAPE, 无法获取有价值的结论^[10]。为便于比较, 给出直接采用原始序列建立GM(1,1)模型预测的2014年各季度发病率数据, 然后分别计算2个模型的精度指标(详见表3)并比较, 发现GM(1,1)模型预测精度指标值均大于前者, 表明GM(1,1)季节指数模型较优。

将2014年发病率数据纳入建模样本, 选择较优的GM(1,1)季节指数模型预测我国2015年各季度菌痢和阿米巴痢疾发病率, 预测值分别为1.4112/10万、3.0674/10万、4.5482/10万、2.0865/10万, 表明2015年我国菌痢和阿米巴痢疾季度发病率继续呈下降趋势, 发病率最高仍在第3季度。

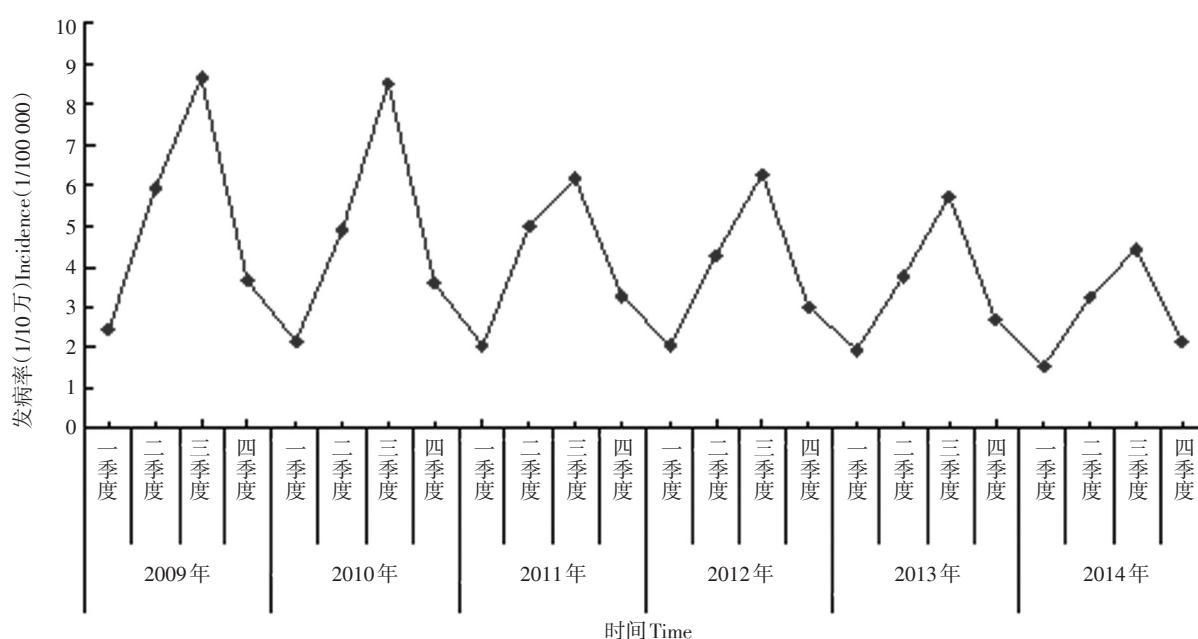


图1 2009—2014年我国菌痢和阿米巴痢疾季度发病率序列

Fig 1 Sequence of quarterly incidence of bacillary dysentery and amebic dysentery in mainland China during 2009–2014

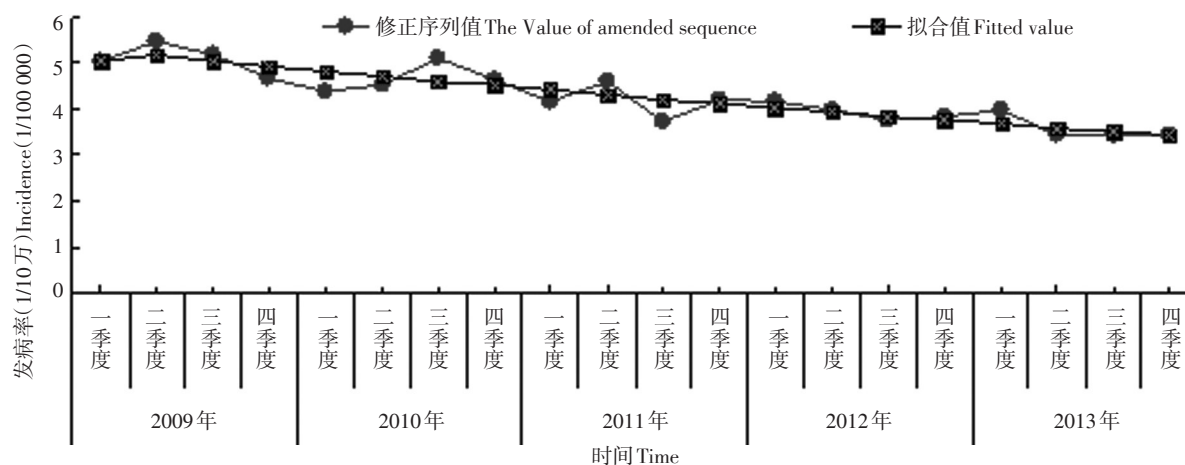


图2 修正序列数据拟合情况

Fig. 2 Fitting illustration of amended sequence data by GM(1,1)

表2 修正序列GM(1,1)模型计算结果

Table 2 Results of GM(1,1)model for amended sequence

<i>t</i>	$x^{(00)}_t$	$x^{(0)}_t$	$x^{(1)}_t$	$\hat{x}^{(0)}_t$
1	2.4539	5.0625	5.0625	5.0625
2	5.9260	5.5034	10.5658	5.1809
3	8.6476	5.2017	15.7676	5.0641
4	3.6389	4.6891	20.4566	4.9500
5	2.1215	4.3766	24.8333	4.8384
6	4.8798	4.5318	29.3650	4.7293
7	8.4945	5.1096	34.4747	4.6227
8	3.5915	4.6281	39.1027	4.5185
9	2.0217	4.1708	43.2735	4.4167
10	4.9552	4.6019	47.8753	4.3171
11	6.1857	3.7208	51.5962	4.2198
12	3.2646	4.2067	55.8029	4.1247
13	2.0191	4.1655	59.9684	4.0317
14	4.2828	3.9773	63.9457	3.9408
15	6.2615	3.7664	67.7122	3.8520
16	2.9846	3.8460	71.5581	3.7651
17	1.9284	3.9784	75.5365	3.6803
18	3.7189	3.4537	78.9902	3.5973
19	5.7324	3.4481	82.4384	3.5162
20	2.6621	3.4304	85.8688	3.4369

3 讨论

由于新发传染病的不断出现和旧的传染病死灰复燃,使得传染病防治仍是我国疾病防控工作的重点^[11],因此基于法定传染病报告系统,收集历史发病数据,积极探索传染病预测方法具有重要的理论价值和现实意义。本次研究以我国菌痢和阿米巴痢疾季度发病率为例,应用GM(1,1)季节指数模型建模预测,并与GM(1,1)模型预测结果比较。结果显示,GM(1,1)季节指数模型预测效果明显优于GM(1,1)模型。

本次研究所涉及的菌痢和阿米巴痢疾季度发病率时间序列资料,发病率总体呈下降趋势,且存在以年为周期的季节性波动,GM(1,1)模型仅能提取其中的线性趋势使得预测效果较差。GM(1,1)季节指数模型,基于非平稳序列的确定性分析原理,将时间序列的季节波动特征用季节指数体现,继而对季节指数修正的去除季节性波动的时间序列进行GM(1,1)模型拟合、预测,最后用季节指数对结果进行还原。因此与GM(1,1)模型相比,GM(1,1)季节指数模型既能客观反映时间序列的总体趋势,也能较好刻画季节性波动特征,并同时保留了不同季节观测值间的相关关系,使得预测精度得以大大改善^[4-5]。

表3 GM(1,1)季节指数模型和GM(1,1)模型预测精度情况

Table 3 Illustration of predictive precision by GM(1,1)season-exponential model and GM(1,1)model

<i>t</i>	实际值 Observed value	GM(1,1)季节指数模型 GM(1,1)season-exponential model			GM(1,1)模型 GM(1,1)model		
		预测值	残差	精度指标*	预测值	残差	精度指标*
		Predictive value	Error	Precision indexes	Predictive value	Error	Precision indexes
21	1.5622	1.6284	-0.0662	0.2505	3.1956	-1.6334	1.3274
22	3.2558	3.5359	-0.2801	0.3889	3.1006	0.1552	0.9968
23	4.4247	5.3360	-0.9113	0.1184	3.0085	1.4162	1.1621
24	2.1368	2.4347	-0.2979	-	2.9191	-0.7823	-

注:精度指标从上到下依次是MSE、MAE、MAPE。Note:The precision indexes were MSE,MAE and MAPE from high to low.

但值得关注的是,运用传染病发病时间序列进行预测时,其预测值大小固然由历史数据决定,但还受传染病发病或流行条件的影响。因此欲提高传染病预测效果,在不断尝试新预测方法的同时,还要搜集传染病影响因素数据,与发病资料同步分析。

参考文献

[1] 万蓉,李娟娟,王晓雯. ARIMA 乘积季节模型在食源性疾病月发病率预测中的应用[J]. 昆明医科大学学报;2012,(6):48- 52.
[2] 许振国,陈双艳,高子厚,等.传染病风险评估体系研究进展[J]. 疾病监测,2012,27(10): 834-838.
[3] 汤云.应用灰色理论预测病毒性肝炎发病情况研究[D].重庆:第三军医大学,2009:18-30.
[4] 刘俊,徐露,黄彦.GM(1,1)季节指数模型在季节性发病情况预测研究中的应用[J].遵义医学院学报,2010,33(4):327-329.

[5] 何俊,张彦群.季节指数修正 GM(1,1)模型在空气质量预测中的应用[J].河南科学,2009,27 (7):779-782.
[6] 张高锋,张永明,黄领梅.灰色季节性指数模型在蒸发预测中的应用[J].水资源与水工程学报,2007,18(6):45-47.
[7] 王琪洁,邹峥嵘,彭悦.混合 GM(1,1)模型预报季节性时间序列精度的方法探讨[J].现代测绘,2003,26(6):11-14.
[8] 程晓宁,庞志明,吴映玫.灰色预测模型广州市白云区水痘防控工作中的应用[J].中国热带医学,2014,14(6):746-748.
[9] 林玫,李永红,黄君,等.GM(1,1)模型在广西四种细菌性传染病流行趋势预测[J].中国热带医学,2011,11(6):671-673.
[10] 杨德平,刘喜华,孙海涛,等.经济预测方法及其 MATLAB 实现[M]. 北京:机械工业出版社,2012:43-44.
[11] 林玫,李永红,董柏青.传染病预测预警方法在我国的应用现状[J]. 中国热带医学,2010,10(3):308-309.

收稿日期:2015-05-08 编辑:谢永慧