

ARIMA和GM(1,1)模型预测上海市卫生总费用

丁海峰, 李立清

湖南农业大学公共管理与法学学院, 湖南 长沙 410128

摘要:卫生总费用是评价一个国家或地区卫生投入水平和资金利用状况的重要指标。文章采用2005—2018年《上海市统计年鉴》相关数据,利用ARIMA模型与灰色GM(1,1)模型,分别对上海市卫生总费用及其占国内生产总值(GDP)比重发展趋势进行预测分析。研究发现,2018—2025年上海市卫生总费用及其占GDP比重将继续保持稳定增长,预测结果显示,2025年上海市卫生总费用及其占GDP比重将分别达到4 765.19亿元(8.53%)、5 976.59亿元(10.21%)。所建立的ARIMA模型与灰色GM(1,1)模型平均相对误差均小于10%,拟合度较好、预测精度较高,但灰色GM(1,1)模型的预测效果要优于ARIMA模型。文章提出今后上海市应在保持卫生总费用继续投入的基础上,控制其合理增长;保持卫生费用的投入与经济发展水平恰当匹配,提高卫生效率,为居民提供更高质量、更加普惠、更有效率的健康服务。

关键词:卫生总费用; ARIMA模型; 灰色GM(1,1)模型; 上海市

中图分类号:C913.4

文献标志码:A

文章编号:1671-0479(2021)05-418-006

doi:10.7655/NYDXBSS20210502

卫生总费用是指一个国家或地区在一定时期内(通常是指1年),全社会用于医疗卫生服务的资金总额^[1]。卫生总费用作为评价一个国家或地区卫生投入水平和资金利用状况的重要指标,了解其未来发展趋势不仅可以为政府部门控制卫生总费用的合理支出,还可以为卫生部门制定医疗卫生政策提供一定的借鉴和参考。正因为卫生总费用的重要程度高,近年来,国内大量学者对卫生总费用的预测进行了探讨。李玲等^[2]利用ARIMA模型对湖南省2018—2022年卫生总费用进行预测分析;辛怡等^[3]通过对我国“十三五”期间卫生总费用占国内生产总值(GDP)的比重进行预测,检验我国卫生总费用占GDP的比重是否达到既定目标;也有学者利用GM(1,1)模型,对上海市卫生总费用及构成进行预测分析^[4]。可以看出,大部分学者均是利用单个模型对我国或者地区的卫生总费用进行分析。然而现实情况中,为提高预测的参考价值,往往需要借助多个模型进行对比分析。基于此,本研究选取上海市2005—2017年卫生总费用及GDP相

关数据,分别建立ARIMA预测模型和灰色GM(1,1)预测模型,对上海市2018—2025年卫生总费用发展趋势进行预测对比分析,一方面可以提高预测的参考价值,另一方面为上海市相关部门制定医疗卫生政策提供借鉴及参考。

一、资料和方法

(一)资料来源

本研究数据来源于2005—2018年《上海市统计年鉴》,其中选取2005—2017年卫生总费用及GDP相关数据作为分析和预测指标,根据两个模型的特点,选取2005—2017年的数据建立ARIMA预测模型,选取2011—2017年的数据建立灰色GM(1,1)模型^[5],数据的来源真实、可靠。

(二)研究方法

1. ARIMA预测模型

ARIMA模型为差分整合移动平均自回归模型,是由学者Box和Jenkins提出的经典时间序列预测方法。在ARIMA(p, d, q)中,AR指自回归,MA指移动平

基金项目:国家社会科学基金后期资助重点项目“我国农村公共产品供给标准理论与实践研究”(20FGLA001);湖南省哲学社会科学成果评审委员会立项资助重大项目“湖南省农村智慧健康养老体系研究”(XSP21ZDA001)

收稿日期:2021-07-26

作者简介:丁海峰(1996—),男,江苏盐城人,博士研究生,研究方向为社会保障、医疗保障。

均, p 指自回归项、 q 指移动平均项数, d 指差分次数。

主要建模步骤包括:①对原始序列进行平稳性检验。可通过序列图观察及单位根检验进行判断,如原始序列呈现非平稳性,则需要对其进行差分处理。绝大多数原始序列均为非平稳序列,均需对其进行差分处理。②模型的识别与检验。差分处理完成后,通过自相关及偏自相关图进行定阶,即确定 p 值和 q 值。如经过差分处理后的平稳序列的自相关图拖尾、偏自相关图截尾时,建立AR模型;当偏自相关图拖尾、自相关图截尾时,则建立MA模型;当自相关和偏自相关图均呈现拖尾,则建立ARIMA模型,通过残差值、AIC等值对比选出最优模型^[6-7]。③参数估计、模型的拟合与结果预测。对所确定的模型进行参数估计,根据所选择的最优模型来检验拟合效果并进行结果预测。

2. 灰色GM(1,1)预测模型

灰色系统理论认为,虽然系统是复杂、杂乱无章的,但系统是一个整体,是有序的。灰色预测就是基于灰色系统的预测,在实际预测中,多采用GM(1,1)模型,即一阶一元灰色模型预测方法,其中“G”代表Grey,表示灰色,“M”代表Model,表示模型,两个“1”分别表示一阶和一个变量的微分方程模型^[8]。灰色预测模型就是通过灰色理论将毫无规律的原始数据进行累加,生成较为规律的数据,从而找出系统内在的特征及逻辑。此方法对样本含量和数据分布没有严格的要求,所需数据单一,原理简单,适用性强,其最大的特点是不需要大量的时间序列数据就能够进行预测,并取得较好的预测效果。

主要步骤包括:①进行级比的检验并对建模的可行性进行分析。②对原数列 $X^{(0)}$ 进行一次累加,AGO(累加生成)为: $x_k^{(1)} = \sum_{i=1}^k x_i^{(0)}, k = 1, 2, \dots, n$ 。③构建数据矩阵 B 及数据向量 Y ,求得预测模型方程式。④对模型预测结果进行检验。

(三)统计学方法

数据的整理与录入在Excel软件中进行,灰色GM(1,1)模型的建立及求解在Matlab 2017软件中进行,ARIMA预测模型的建立及求解在SPSS24.0统计软件中进行。通过描述性统计法对上海市2005—2017年卫生总费用发展现状及趋势进行对比分析。分别利用ARIMA预测模型与灰色GM(1,1)预测模型对上海市2018—2025年卫生总费用及其占GDP比重进行预测对比分析。

二、结 果

(一)上海市卫生总费用及其占GDP比重发展现状

由表1可知,2005—2017年,上海市卫生总费用及其占GDP的比重均呈现不断增长的趋势,卫生总

费用从362.13亿元增长至2 087.00亿元,年平均增长率为15.7%;卫生总费用占GDP的比重从3.87%增长至6.81%,年平均增长率为4.83%。

表1 2005—2017年上海市卫生总费用及占GDP比重

年份	卫生总费用(亿元)	占GDP比重(%)
2005	362.13	3.87
2006	401.46	3.75
2007	485.67	3.83
2008	559.83	3.92
2009	656.66	4.30
2010	751.99	4.31
2011	931.00	4.76
2012	1 092.35	5.31
2013	1 248.68	5.61
2014	1 347.79	5.60
2015	1 536.60	5.99
2016	1 838.00	6.52
2017	2 087.00	6.81
年平均增长率(%)	15.70	4.83

(二)ARIMA模型的预测过程及结果

1. 数列的平稳性处理

根据SPSS24.0软件所绘制的2005—2017年上海市卫生总费用及占GDP比重的序列图可以看出,两个指标均呈现明显增长趋势,均不是平稳序列。因此,需对其进行差分处理。经过检验分析后可知,上海市卫生总费用序列在经过2阶差分后趋于平稳,故 $d=2$;卫生总费用占GDP比重序列经过1阶差分后趋于平稳,故 $d=1$ 。原始序列及差分序列图见图1、2。

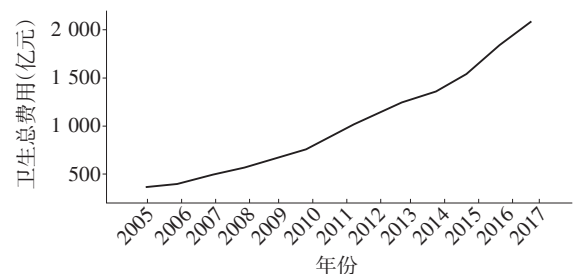


图1 上海市卫生总费用原始序列图

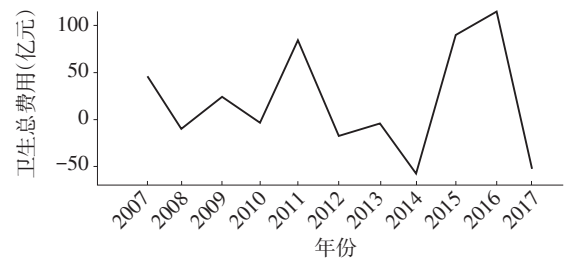


图2 上海市卫生总费用2阶差分图

2. 参数的估计及检验

差分处理之后,利用SPSS24.0软件对上海市卫生总费用及占GDP比重分别进行拟合,通过自相关及偏自相关图初步判断拖尾和截尾情况(图3、4),经过模型的对比和不断尝试,估计ARIMA模型的各项参数,最终确定卫生总费用预测模型最优的3个

参数为 $p=0, d=2, q=0$; 卫生总费用占 GDP 比重预测模型最优的 3 个参数为 $p=0, d=1, q=1$ 。分别建立 ARIMA(0, 2, 0) 及 ARIMA(0, 1, 1) 预测模型。经检验可知, 模型的拟合统计量 R^2 分别为 0.988、0.955, R^2 均 >0.9 ; 标准化贝叶斯信息量 (BIC) 值分别为 8.311、-2.761; 平均绝对误差 (MAE) 分别为 46.948、0.196; 残差的 P 值均 >0.05 。由此可以看出模型的拟合效果较好, 残差为白噪声序列, 可以利用其对上海市卫生总费用及其占 GDP 比重进行预测。

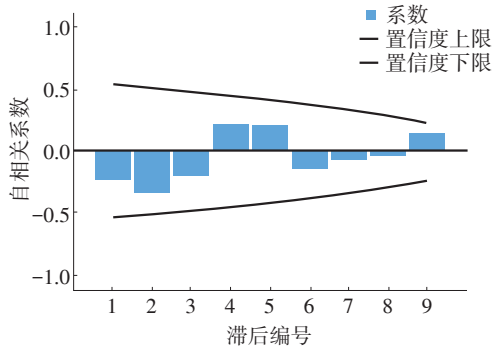


图3 上海市卫生费用二阶差分自相关

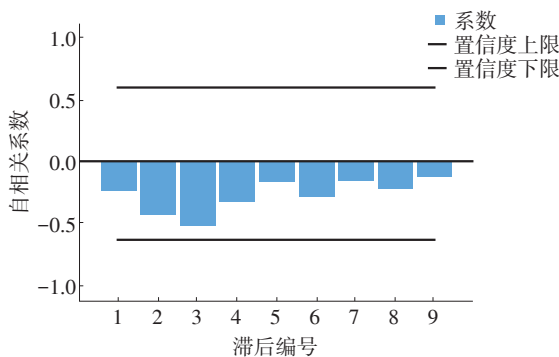


图4 上海市卫生费用二阶差分偏自相关

3. 模型的拟合及预测

利用所建立的 ARIMA 模型分别对上海市卫生总费用及其占 GDP 的比重进行预测, 并求出相对误差 [相对误差 = (预测值 - 实际值) / 实际值]。经计算, 上海市实际值与预测值平均相对误差绝对值仅为 0.74%, 说明拟合效果较好, 可以利用其进行外推预

测。卫生总费用预测值与实际值对比情况见表 2。

(三) 灰色 GM(1, 1) 模型预测过程及结果

① 建立时间序列:

$$X^{(0)} = [x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(7)] = (931.00, 1\ 092.35, \dots, 2\ 087.00)$$

经过计算, 该数列所有的级比数值都落在区间

$$\sigma(k) = \frac{x^{(0)}(k-1)}{x^{(0)}(k)} \text{ 内, 即区间 } (0.778\ 800\ 783, 1.284\ 025\ 417)$$

内, 可以进行模型的建立。

② 通过累加构造生成列: (931.00, 2 023.35, 3 272.03, 4 619.82, 6 156.42, 7 994.42, 1 0081.42)。

③ 构造数据矩阵 B 和数据向量 Y :

$$B = \begin{pmatrix} -\frac{1}{2}[x^{(0)}(1) + x^{(0)}(2)] & 1 \\ -\frac{1}{2}[x^{(0)}(2) + x^{(0)}(3)] & 1 \\ \vdots & \vdots \\ -\frac{1}{2}[x^{(0)}(6) + x^{(0)}(7)] & 1 \end{pmatrix}$$

$$Y = \begin{pmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(7) \end{pmatrix}$$

经计算 $B^T B$, $(B^T B)^{-1}$, $\hat{a} = (B^T B)^{-1} \times B^T \times Y$, 得到

$\frac{\mu}{a} = -6\ 553.809\ 9$, 最终得到上海市卫生总费用的预测模型为 $\hat{X}^{(1)}(k+1) = 870.455\ 303\ 3e^{0.132\ 816\ 684k} - 6\ 553.809\ 9$

同理, 上海市卫生总费用占 GDP 的比重序列也通过级比检验, 因此, 可以建立 GM(1, 1) 灰色预测模型。故可以计算出卫生总费用占 GDP 比重的 a 和 μ 分别为: -0.051 446 433、0.048 524 838。因此, 可以建立上海市卫生总费用占 GDP 比重的 GM(1, 1) 灰色预测模为:

$$\hat{X}^{(1)}(k+1) = 0.048\ 524\ 838e^{0.051\ 446\ 433k} - 0.943\ 210\ 9$$

④ 模型的检验。经计算可知, 上海市卫生总

表2 上海市卫生总费用及其占 GDP 比重实际值与预测值对比 (ARIMA 模型)

年份	卫生总费用(亿元)			占 GDP 比重(%)		
	实际值	预测值	相对误差	实际值	预测值	相对误差
2005	362.13	—	—	3.87	—	—
2006	401.46	—	—	3.75	—	—
2007	485.67	459.85	-0.05	3.83	4.12	0.08
2008	559.83	588.94	0.05	3.92	4.00	0.02
2009	656.66	653.05	-0.01	4.30	4.08	-0.05
2010	751.99	772.55	0.03	4.31	4.17	-0.03
2011	931.00	866.38	-0.07	4.76	4.55	-0.04
2012	1 092.35	1 129.07	0.03	5.31	4.56	-0.14
2013	1 248.68	1 272.76	0.02	5.61	5.01	-0.11
2014	1 347.79	1 424.07	0.06	5.60	5.56	-0.01
2015	1 536.60	1 465.96	-0.05	5.99	5.86	-0.02
2016	1 838.00	1 744.47	-0.05	6.52	5.85	-0.10
2017	2 087.00	2 158.46	0.03	6.81	6.42	-0.06

费用的平均相对误差为2.34%,后验差比值 C 为0.006 896 762, C 远小于0.35,模型精度 P 为97.66%,表明模型的精度较高,同理,求得卫生总费用占GDP比重预测模型的后验差比值为:0.027 260 739,模型精度 P 为98.17%。经模型验证,所有 C 均小于0.35,

由此可以表明所建立的模型精度相对较好,均具有较为科学、准确的预测效果。由结果可知,所有模型 $-a$ 均 <0.3 ,因此可以利用所构建的模型对上海市卫生总费用及其占GDP比重进行中长期的预测,实际值与预测值的结果见表3,具体拟合效果见图5、6。

表3 上海市卫生总费用及其占GDP比重实际值与预测值对比[GM(1,1)模型]

年份	卫生总费用(亿元)			占GDP比重(%)		
	实际值	预测值	相对误差	实际值	预测值	相对误差
2011	931.00	—	—	4.76	—	—
2012	1 092.35	1 063.15	0.03	5.31	5.23	0.02
2013	1 248.68	1 214.16	0.03	5.61	5.51	0.02
2014	1 347.79	1 386.62	-0.03	5.60	5.80	-0.04
2015	1 536.60	1 583.57	-0.03	5.99	6.10	-0.02
2016	1 838.00	1 808.50	0.02	6.52	6.43	0.01
2017	2 087.00	2 065.39	0.01	6.81	6.77	0.01

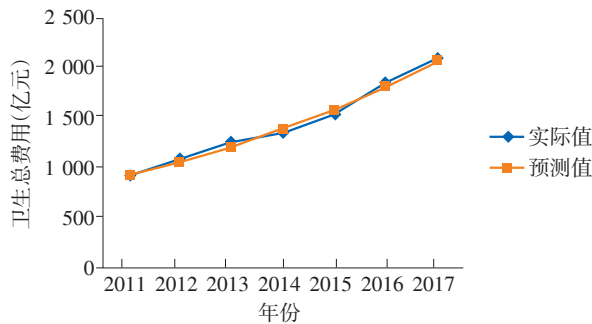


图5 上海市卫生总费用灰色GM(1,1)预测模型拟合情况

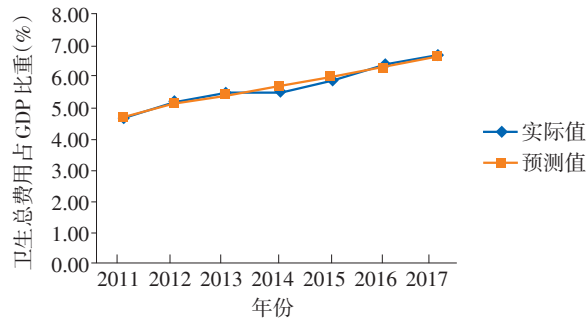


图6 上海市卫生总费用占GDP比重灰色GM(1,1)预测模型拟合情况

(四) ARIMA模型与灰色GM(1,1)模型预测结果及误差对比

由表4可以看出,2018—2025年,上海市卫生总费用及占GDP比重将保持稳步上升的趋势。ARIMA预测模型结果显示,到2025年,上海市卫生总费用将达到4 765.19亿元,卫生总费用占GDP的比重将达到8.53%;灰色GM(1,1)预测模型的结果显示,到2025年,上海市卫生总费用将达到5 976.59亿元,卫生总费用占GDP的比重将达到10.21%。可以看出两种预测模型所预测的结果存在一定的差距。

表4 ARIMA模型与灰色GM(1,1)模型预测结果

年份	卫生总费用(亿元)		卫生总费用占GDP比重(%)	
	ARIMA模型	GM(1,1)模型	ARIMA模型	GM(1,1)模型
2018	2 355.06	2 358.75	6.77	7.12
2019	2 642.18	2 693.79	7.06	7.50
2020	2 948.37	3 076.42	7.30	7.89
2021	3 273.61	3 513.40	7.55	8.31
2022	3 617.91	4 012.44	7.79	8.75
2023	3 981.28	4 528.37	8.04	9.21
2024	4 363.71	5 233.26	8.28	9.70
2025	4 765.19	5 976.59	8.53	10.21

ARIMA预测模型和灰色GM(1,1)模型的平均相对误差存在一定的差距。在卫生总费用方面,GM(1,1)预测模型的平均相对误差为0.03亿元,ARIMA模型的平均相对误差为0.04亿元,GM(1,1)预测模型要小于ARIMA模型。在卫生总费用占GDP比重方面,GM(1,1)模型的平均相对误差为0.02%,而ARIMA模型的平均相对误差达到了0.07%。由此可见,对于上海市卫生总费用及其占GDP的比重进行预测,灰色GM(1,1)预测模型的预测精度及效果要优于ARIMA模型。

三、讨论

(一) 未来上海市卫生总费用及其占GDP比重将保持稳步增长

预测结果显示,2018—2025年上海市卫生总费用及其占GDP比重将继续保持稳步增长趋势。卫生总费用的投入水平不仅可以反映一个国家和地区医疗卫生水平的高低,还可以反映当地政府对居民健康水平的重视程度^[9]。近年来,上海市不断加

大卫生总费用的投入,这一方面表明上海市具有较高的医疗卫生水平,另一方面则表明当地政府对居民健康的重视。然而,卫生总费用的过快增长不仅不利于当地医疗卫生水平的提高,而且会对政府财政造成极大的负担^[10]。因此,上海市相关部门在保持卫生总费用稳步增长的基础上,应该控制卫生总费用的合理增长,同时保证政府卫生支出、社会卫生支出以及个人卫生支出在合理的范围内,不断优化卫生筹资结构。此外,卫生总费用占GDP的比重也可以反映一个国家或地区医疗卫生水平的高低。根据ARIMA模型的预测结果,到2025年,上海市卫生总费用占比将达到8.53%,而灰色GM(1,1)模型预测结果显示,这一指标将达到10.21%。按照世界卫生组织的要求,发展中国家卫生总费用占GDP总费用不应低于5%。因此,无论如何,两个模型的预测值都已经超过世界卫生组织规定的最低标准,其实早在2012年上海市就已经达到这一标准。虽然世界各国关于卫生总费用占GDP比重这一指标的标准尚未统一,但至少反映卫生费用的投入水平和国民经济发展水平的相互匹配程度。因此,在今后的工作中,上海市要保持卫生费用的投入与经济发展水平恰当匹配,提高卫生效率,为居民提供更高质量、更加普惠、更有效率的健康服务。

(二)灰色GM(1,1)模型预测效果优于ARIMA模型

本研究选取2005—2017年上海市卫生总费用及GDP相关数据,利用ARIMA模型和灰色GM(1,1)模型分别对上海市2018—2025年卫生总费用及其占GDP的比重进行预测分析。结果显示,两个模型的平均相对误差均低于10%,说明所建立的预测模型精度较高,均呈现较好的预测效果。但灰色GM(1,1)模型的平均相对误差略低于ARIMA模型,这与部分学者的研究结果不一致^[11],可能与本文预测的对象以及数据周期的选取有一定的关系。根据卫生总费用预测结果,ARIMA模型的平均相对误差为0.04亿元,GM(1,1)模型的平均相对误差为0.03亿元,可以看出灰色GM(1,1)模型的预测精度略高于ARIMA模型;而对于卫生总费用占GDP比重的预测上,两者差距较大,ARIMA模型的平均相对误差为0.07%,而灰色GM(1,1)模型的平均相对误差仅为0.02%。灰色GM(1,1)模型和ARIMA模型是医疗卫生领域常用的预测模型,在短期预测中均有较好的预测效果^[12]。然而,在现实研究中,不同预测对象的特征、属性等均

有较大的差别,再加上两种模型预测原理存在一定的区别,这就导致两个模型会呈现截然不同的预测效果。因此,在今后的预测研究中,要根据不同预测对象的特点及属性,选择合适的预测模型进行分析,同时可以采用多种预测模型进行组合分析,这样不仅可以提高预测的效果和精度,还可以为决策者制定相关政策提供多种参考依据。

参考文献

- [1] 孟庆跃. 卫生经济学[M]. 北京:人民卫生电子音像出版社,2013:7-8
- [2] 李玲,夏新斌,周良荣. 基于ARIMA模型的湖南省卫生总费用趋势预测及分析[J]. 湖南农业大学学报(社会科学版),2019,20(5):69-75
- [3] 辛怡,马蔚姝,李惠,等. “十三五”时期我国卫生总费用占国内生产总值比重预测[J]. 现代预防医学,2018,45(10):1804-1808
- [4] 丁海峰,高凯,姜茂敏. 基于灰色GM(1,1)模型的上海市卫生总费用预测研究[J]. 医学与社会,2020,33(6):42-46
- [5] 潘静,张颖,刘璐. 基于ARIMA模型与GM(1,1)模型的居民消费价格指数预测对比分析[J]. 统计与决策,2017(20):110-112
- [6] 孙颖. 基于ARIMA模型的消费者价格指数预测[J]. 统计与决策,2016(11):83-85
- [7] 孙娜,许小珊,冯佳宁,等. ARIMA与GM(1,1)模型对我国肺结核年发病人数预测情况的比较[J]. 中国卫生统计,2019,36(1):71-74
- [8] 朱春梅. 灰色GM(1,1)模型在孕产妇和新生儿死亡率预测研究中的应用[J]. 中国妇幼保健,2016,31(3):455-457
- [9] 李燕凌,李立清. 基于ARIMA模型的中国卫生总费用预测分析[J]. 中国卫生事业管理,2010,27(8):508-510,520
- [10] 王高玲,张怡青. 基于ARIMA模型的我国卫生总费用趋势及构成预测分析[J]. 中国卫生统计,2019,36(1):121-125
- [11] 张瑞华,赵大仁,何思长,等. 基于灰色GM(1,1)模型与ARIMA模型的四川省卫生人力资源预测探讨[J]. 现代预防医学,2017,44(7):1242-1247
- [12] 胡跃华,廖家强,冯国双,等. ARIMA模型在全国丙型肝炎疫情预测中的应用[J]. 中国预防医学杂志,2015,16(4):262-266

(本文编辑:姜鑫)

Forecast analysis of Shanghai's total health expenditure based on ARIMA and GM(1,1) models

DING Haifeng, LI Liqing

School of Public Administration and Law, Hunan Agricultural University, Changsha 410128, China

Abstract: Total health expenditure is an important index to evaluate the health investment level and fund utilization status of a country or region. Based on the data of Shanghai Statistical Yearbook from 2005 to 2018, ARIMA model and grey GM(1,1) model were used to forecast and analyze the development trend of total health expenditure and its proportion in GDP in Shanghai. The study found that the total health expenditure and its proportion in GDP of Shanghai would continue to grow steadily from 2018 to 2025. The forecast results showed that the total health expenditure and its proportion in GDP of Shanghai would reach 476.519 billion yuan, 597.659 billion yuan, 8.53% and 10.21% respectively in 2025. The average relative errors of the established ARIMA model and the grey GM(1,1) model were both less than 10%, and the fitting degree was good and the prediction accuracy was high, but the grey GM(1,1) model was better than the ARIMA model. The paper puts forward that Shanghai should control the reasonable increase of the total health expenditure on the basis of keeping the investment, maintain proper matching between health expenditure and the level of economic development, improve health efficiency, and provide people with higher quality, more inclusive and more efficient health services.

Key words: total health expenditure; ARIMA model; grey GM(1,1) model; Shanghai