维普资讯 http://www.cqvip.com

论 著 文章编号:0253-3626(2008)04-0475-03

ARIMA 季节乘积模型在预测医院月住院量的应用

刘金莲,钟晓妮

(重庆医科大学公共卫生学院卫生统计教研室,重庆 400016)

【摘 要】目的:探讨 ARIMA 季节乘积模型在时间序列资料分析中的应用,建立住院量的预测模型。方法:利用重庆市某医院 2002~2005 年的月住院量,采用条件最小二乘法估计模型参数,按照残差不相关原则、简洁原则确定模型结构,依据 AIC 和 SBC 准则确定模型阶数,建立 ARIMA 预测模型。结果:模型参数有统计学意义。方差估计值为 0.003 175,AIC=-94.245 1,SBC=-91.252 1。对模型进行白噪声残差分析,拟合优度结果表明 ARIMA 的最优模型为:$(1-B)(1-B^{12})Z_t=(1-0.543\ 02B)(1-B^{12})a_t$。结论:用所建立模型对月住院量进行预测,结果表明 ARIMA 是一种短期预测精度较高的预测模型。

【关键词】时间序列;预测;自回归求和移动平均模型

【中国图书分类法分类号】R197.323 【文献标识码】A 【收稿日期】2007-09-05

Applications of multiple seasonal ARIMA model on predictive workload of inpatient department

LIU Jin-lian, et al

(*Department of Statistics, College of Public Health, Chongqing Medical University*)

【Abstract】*Objective*: To discuss the application of multiple seasonal ARIMA model to data in time series and fit predictive model on workload of inpatient department. *Methods*: Parameter of model is estimated using conditional least squares method according to the data of workload of inpatient department in a hospital in Chongqing from 2002 to 2005. The structure is determined according to criteria of residual un-correlation and concision. ARIMA predictive model was fitted and the order of model was confirmed through Akaike's Information Criterion and Schwarz Bayesian Criterion. *Results*: Statistics assisted estimation of the model. The estimation of variance is 0.003 175, AIC=-94.245 1, SBC=-91.252 1. Using analysis of white-noise residual of model, fit of rich table show that the best estimated ARIMA model is $(1-B)(1-B^{12})Z_t=(1-0.543\ 02B)(1-B^{12})a_t$. *Conclusion*: The model of ARIMA can be used to forecast for workload of inpatient department with high prediction precision of short-term time series.

【Key words】 Time series; Prediction; ARIMA model

统计预测是医院科学决策的依据,对医院住院量进行短期预测,对于合理安排人、财、物资源,提高经济效益和社会效益都有重要意义。时序模型是常用的预测方法。建立时序模型的方法众多,有移动平均法、趋势预测模型、周期趋势预测模型、季节趋势预测模型、指数平滑模型、谱分析法[1-3]等。本研究利用重庆市某医院 2002 年 1 月 ~2005 年 10 月的住院量,探讨 ARIMA 季节乘积模型、指数平滑模型的短期预测效果,为合理配置卫生资源和平衡卫生服务需求提供科学依据。

作者介绍:刘金莲(1982-),女,硕士,
研究方向:大气环境对健康影响评价。
通讯作者:钟晓妮,女,副教授,E-mail:zxn66@vip.sina.com。

1 资料与方法

1.1 资料来源

采集了重庆市某医院 2002 年 1 月到 2005 年 10 月的入院病人信息,整理为逐月住院量。

1.2 ARIMA 模型

ARIMA 模型即自回归 - 求和 - 移动平均模型,由 Box 与 Jenkins 于 1976 年提出,指序列 Zt 的 d 阶差分具有 ARMA(P,q)模型,记做 ARIMA(P,d,q)。其模型为:$Z_t=\phi_1 Z_{t-1}+\phi_2 Z_{t-2}+\cdots\cdots+\phi_p Z_{t-p}\phi_1+e_t-\theta_1 e_{t-1}-\theta_2 e_{t-2}-\cdots\cdots-\theta_q e_{t-q}$,其中 p 是自回归模型的阶数,$\phi_i$ 为自回归系数,Z_{t-i} 是时间序列在 t-i 期的观察值,e_t 是不能用模型说明的随机扰动;q 是移动平均模型的阶数,θ_2 为移动平均系数,e_{t-i} 为第 t-i 个时期进入系统的随机扰动。其应用的前提条件是:时间序列应是一

维普资讯 http://www.cqvip.com

零均值的平稳随机序列。

有的时间序列不仅含有季节性成分,还混有非季节性成分,若单一用季节性或非季节性 ARIMA 模型进行分析,其预测效果往往不理想。季节性 ARIMA 与非季节性 ARIMA 的混合效应通常表现为相乘的形式,即乘积 ARIMA(P,d,q)(P,D,Q)s 模型。ARIMA 季节乘积模型可以综合考虑序列演变的趋势、周期变化和随机干扰因素,借助模型参数的变化对数据进行量化表达,可以达到较好的预测效果。

2 结 果

2.1 资料特征

2002 年 1 月 ~2005 年 10 月,该医院的逐月住院量见图 1。从时序图可以看出:数据有明显的上升趋势,且年内也有季节周期性变化,1、2 月的住院人次最少,与春节有一定关系,在 2003 年 5 月也有个低值,非典的流行应该是其主要原因。

2.2 时间序列模型的建立步骤

ARIMA 建模法分为模型识别,参数估计和诊断检验 3 个阶段,往往需要 3 个过程反复进行,最终确定 1 个用于预报或控制的最优模型。

2.2.1 模型识别 在对原始数据进行对数转换的基础上进行了一阶非季节性差分和一阶周期为 12 的季节性差分,将序列平稳化,白噪声检验如表 1 所示,说明平稳后的序列有相关关系,为非随机序列,可以拟合 ARIMA 模型[4]。由序列的自相关系数和偏自相关系数初步判断模型为 ARIMA(0,1,1)(0,1,1)12。

2.2.2 参数估计 非季节和季节移动平均参数分

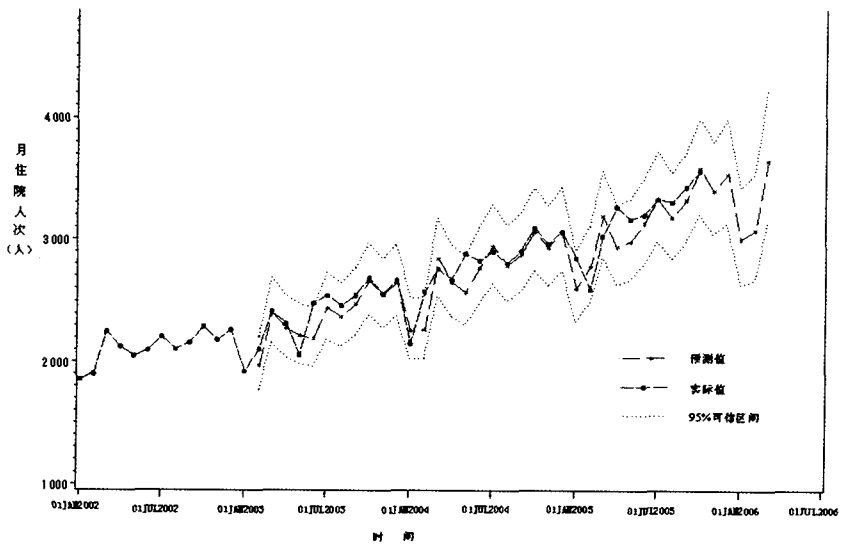


图 1 2002 年 1 月~2005 年 10 月住院人数时序图

表 1 时间序列的白噪声检验结果

延迟阶数	χ^2 值	自由度	P 值	自相关系数					
6	14.08	6	0.028 8	−0.533	−0.078	0.241	−0.170	0.070	−0.004
12	46.98	12	0.000 1	−0.169	0.435	−0.338	−0.066	0.398	−0.401
18	48.05	18	0.000 1	0.122	0.003	0.023	0.015	−0.028	−0.037
24	56.15	24	0.000 2	0.132	−0.230	0.136	0.017	−0.031	−0.018
30	56.85	30	0.002 2	−0.005	0.050	−0.035	−0.010	0.009	0.004

表 2 模型参数估计

模型参数	估计值	标准误	t 值	P 值	延迟
MA1,1	0.543 02	0.179 85	3.02	0.005 0	1
MA2,1	1.000 00	0.211 89	4.72	0.000 1	12

维普资讯 http://www.cqvip.com

表 3 模型对未来 5 月的预测结果

时间	预测值	95%可信区间的上限	95%可信区间的下限
2005 年 11 月	3 401.12	3 401.12	3 040.68
2005 年 12 月	3 536.41	3 536.41	3 126.05
2006 年 1 月	3 000.78	3 000.78	2 625.13
2006 年 2 月	3 077.79	3 077.79	2 666.55
2006 年 3 月	3 641.83	3 641.83	3 126.64

别为 0.543 02 和 1.000 00，t 检验的 P 值均小于 0.05，有统计学意义，方差估计值为 0.003 175，AIC=-94.245 1，SBC=-91.252 1，显示模型提取序列中几乎所有的样本相关信息。

2.2.3 模型诊断 对残差序列进行白噪声检验：P=0.440 7，说明为白噪声序列，模型 $(1-B)(1-B^{12})Z_t=(1-0.543\ 02B)(1-B^{12})a_t$ 具有较佳的拟合优度。

2.2.4 模型预测 根据建立的 ARIMA(0,1,1)(0,1,1)12 预测模型，对未来 5 个月的住院量进行了短期预测，其点预测值分别为 3 401.12、3 536.41、3 000.78、3;077.79、3 641.83;95%可信区间(3 040.68,3 792.2)、(3 126.05,3 985.31)、(2 625.13,3 414.78)、(2 666.55,3 534.15)、(3 126.64,4 217.26)。序列的预测值及 95%可信区间见图 1 所示，预测的相对误差在 0.13%和 11.90%之间，平均值为 3.80%，实际值均在预测值的可信区间范围，可见模型的预测精度较高。

3 比 较

为了比较不同模型的效果，尝试拟合指数平滑模型，模型的常规参数 α=0.3，平均相对误差为 5.63%，可见 ARIMA 模型的拟合效果优于指数平滑模型。

4 讨 论

时序分析克服了因果回归分析法中预测对象的影响因素难以掌握和数据资料不易得到的难题，利用任何事物的发展都具有一定惯性(即延续性)的原理，建立时间序列模型，以达到预测未来的目的[5]。它广泛用于各个领域，在医学及卫生事业管理的应用主要体现在：医院统计和医院管理、疾病流行趋势的分析和预测、环境医学领域等[6]。

建立时序模型的方法众多，大多数传统的时序模型均假设各变量之间是一种线性关系，其过程简便实用，但是对于有季节性变化趋势的病人就诊量资料的预测效果往往不佳。本研究根据资料特征选择了ARIMA季节乘积模型，将原始数据取对数后，通过一阶季节性差分和一阶非季节性差分使原序列平稳化，并建立了最优模型。用该模型预测的平均相对误差为 3.80%，低于指数平滑模型，可见 ARIMA 季节模型优于指数平滑模型，这与孙木、张蔚等的研究结果相同[2,7]。综上，ARIMA 季节乘积模型对医院月住院量的短期预测较好，对卫生资源的安排以及医院的管理提供了科学依据。

参 考 文 献

[1] 于亚滨.医院门急诊、住院病人量预测及变动趋势分析[J]. 中国病案,2005;6(6):43-45.

[2] 孙 木,袁克俭,孙海琴.预测模型在门诊量分析与预测中的应用[J]. 中国医院统计,2006;13(4):343-345.

[3] 覃世龙,杨 飞,司红玉,等.谱分析方法在医院时间序列资料分析中的应用[J]. 中国医院统计,2006;13(3):243-245.

[4] 王 燕,编著.应用时间序列分析[M]. 中国人民大学出版社,2005.

[5] 吴家兵,叶临湘,尤尔科.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.

[6] 邓 丹,王润华,周燕荣.时间序列分析及其在卫生事业中的应用[J]. 数理医药学杂志,2002;15(5):455-457.

[7] 张 蔚,张彦琦,杨 旭.时间序列资料 ARIMA 季节乘积模型及其应用[J]. 第三军医大学学报,2002;24(8):955-957.

(责任编辑：冉明会)