

探讨 ARIMA 模型在红细胞用量预测中的应用*

袁玉荣¹ 高均翠¹ 卢伟¹ 张莉¹ 郭毅¹ 郑山根²

[摘要] **目的:**分析宜昌地区 ABO 血型系统各血型红细胞用量的分布规律,依据时间序列分析方法建立自回归积分滑动平均模型(ARIMA)进行预测,指导采供血机构相关业务工作。**方法:**在 SPSS18.0 中利用时间序列模型中专家建模器,对宜昌市 2008-01—2015-12 红细胞总的用量及各血型分别用量建立数学模型,并预测 2016 年 1 至 6 月用量,与实际用量对比,验证模型误差。**结果:**专家建模器对红细胞总量、A 型和 O 型红细胞用量给出的模型是 ARIMA(0,1,1)(0,1,1),B 型和 AB 型红细胞用量给出的模型分别是 ARIMA(1,1,1)(1,1,1)和 ARIMA(2,1,1)(1,1,1)。对 5 个模型残差的白噪声检验结果均显示 $P>0.05$,说明残差均为白噪声序列,模型提取了原序列中所有数据信息,模型诊断得以通过。将预测值与实际值进行比较,实际值均落入预测值 95% 的可信区间内,且平均误差相对较小,模型预测效果良好。**结论:**ARIMA 模型能够科学、有效地反映时间序列的变化规律,可以有效预测短期红细胞用量,有针对性地指导血站的采供血业务工作。

[关键词] ARIMA 模型; 悬浮红细胞; 用量

doi:10.13201/j.issn.1004-2806-b.2017.08.006

[中图分类号] R555 **[文献标志码]** A

Discussion on application of ARIMA model in the prediction of red cell use volume

YUAN Yurong¹ GAO Juncui¹ LU Wei¹ ZHANG Li¹
GUO Yi¹ ZHENG Shangen²

(¹The Center Blood Bank of Yichang City, Yichang, 443005, China; ²Department of Clinical Transfusion, Wuhan General Hospital of Guangzhou Military Command)

Abstract **Objective:** To analyze the distribution of the red blood cell use volume of ABO blood group in Yichang area, and establish the ARIMA model based on time series analysis method to predict the volume and guide the relevant works of the blood collection and supply organizations. **Method:** We used time series expert modeler in SPSS18.0 to set up mathematical models for the monthly total use volume and each kind use volume of red blood cells from January 2008 to December 2015 in Yichang city, and used them to predict the monthly use volume from January to June 2016 and compared them with the actual volume to verify the model error. **Result:** The expert modeler gave ARIMA(0,1,1)(0,1,1) for the total use volume and A and O type red blood cell use volume, ARIMA(1,1,1)(1,1,1) for B type red blood cell use volume, and ARIMA(2,1,1)(1,1,1) for AB type red blood cell use volume. The residual white noise test results of the five models were all $P>0.05$, indicating these residuals were white noise and the models extract all the data from the original sequence and passed the test. Comparing the predicted results with the actual values, all of the actual values fell into the predicted interval of 95% confidence level, and their relative errors were small. The predictive effect of the models was perfect. **Conclusion:** The ARIMA model can scientifically and effectively reflect the changes of time series, effectively predict the short-term use volume of red blood cells, and pertinently guide the blood collection and supply work of blood bank.

Key words ARIMA model; suspension red blood cell; blood use volume

血液是宝贵的稀缺资源,目前血液保存的技术下,采集的新鲜血液最多只能保存 35 d,血站作为采供血机构储存血液过多容易导致血液报废,过少不能满足医疗机构的用血需求,甚至影响急危重症患者的生命。如何保证采供血机构采血和供血之间的平衡,按临床需求有目的有计划地控制采血量,既能保证有效临床供血,又减少血液偏型与浪

费,是血站质量管理中的一个重要问题。

本研究采用自回归积分滑动平均模型(autoregressive integrated moving average, ARIMA),对我站 2008-01—2015-12 悬浮红细胞用量进行统计分析,建立预测模型,为本站合理备血提供参考和指导。ARIMA 是由博克思和詹金斯于 70 年代初提出的一著名时间序列预测方法,指将非平稳时间序列转化为平稳时间序列,然后将因变量仅对它的滞后值以及随机误差项的现值和滞后值进行回归所建立的模型。

* 基金项目:湖北省卫生和计划生育委员会科研项目(No: WJ2015CB008);宜昌市科技局科研项目(No:A1530142)

¹宜昌市红十字中心血站(湖北宜昌,443005)

²广州军区武汉总医院

1 资料与方法

1.1 资料来源

以我站 2008-01—2015-12 每月各型红细胞的用量为原始数据。以月为时间单位进行统计用量, 用量以单位(U)计算, 悬浮红细胞 1 U 为 200 ml。数据来源为我站信息科数据库。

1.2 方法

利用 Excel 建立数据库, 按月统计我站用血总量, 并分 A、B、O、AB 4 种血型统计红细胞的用量。红细胞用血总量及各型红细胞用量不确定, 但随着时间的和季节的变化有一定的规律性。本研究采用 ARIMA 模型, 分别建立用血总量及各型红细胞的用量的预测模型。

ARIMA 模型建立后, 将数据导入 SPSS18.0 建立红细胞数据库, 通过 4 个步骤序列平稳化、模型识别、模型参数的估计与检验及预测应用等来建立合适的预测模型进行预测应用^[1-2]。

2 结果

2.1 红细胞月供应总量的时间序列特性

我站红细胞供应具有波动性, 且表现出一定的季节性, 每年 1~2 月是用血量的低谷, 而 10~12 月是用血量的高峰时期, 且每年的用水量逐年增加, 2013—2015 年逐渐平稳, 上升趋势减缓。

2.2 建模 经 SPSS18.0 软件建模

通过专家建模器, 得到每个血型用量及用血总量较好的模型(表 1、图 1~4), 由表 1 中各模型的相关参数平稳的 R 方、MAPE(平均绝对误差百分比)、MAE(平均绝对误差)、MaxAE(最大绝对误差)及正态化的 BIC 看出, 模型残差的白噪声结果均为 $P>0.05$, 显示残差均为白噪声序列, 各模型提取了原序列中的原始数据, 模型诊断具有较好的适合性。

2.3 数据预测分析

各模型对 2016 年 1 至 6 月份的数据进行预测分析后, 显示 5 个模型红细胞类每月的实际值均落入 95% 的可信区间内, 相对误差均 $<10\%$, 平均相对误差较小, 见表 2。

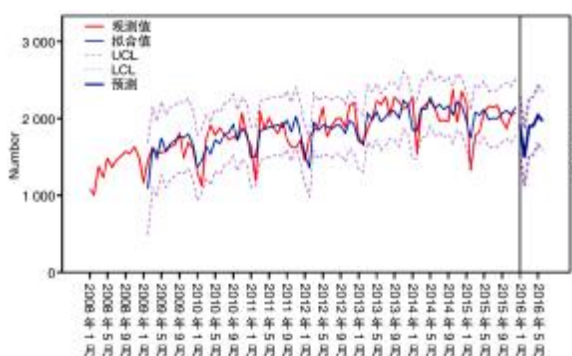


图 1 A 型红细胞月用量模型

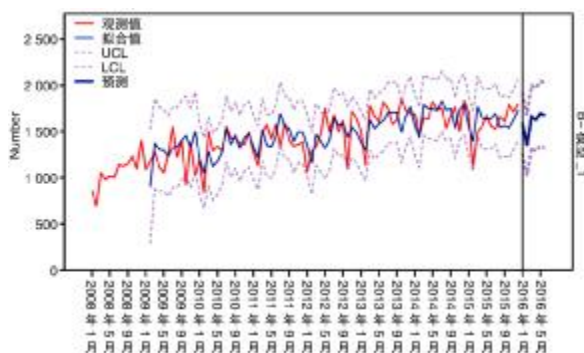


图 2 B 型红细胞月用量模型

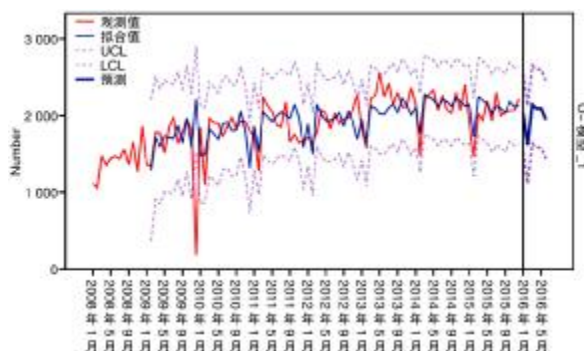


图 3 O 型红细胞月用量模型

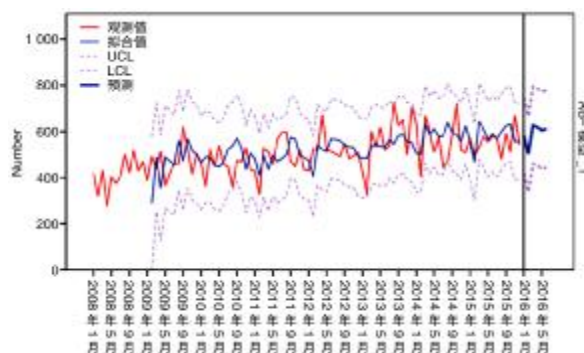


图 4 AB 型红细胞月用量模型

3 讨论

近年来我市经济社会发展迅速, 随着分级诊疗的模式不断完善, 医疗需求将会不断增加, 安全高效供血显得尤为重要。我国部分地区做了无偿献血人群的血型分布调查, 但临床用血是否与无偿献血人群的血型分布有一致性, 未见报道。无偿献血容易受天气、季节及高校学生放假等因素的影响, 采血量不稳定, 一般情况下只能根据库存量来初步调整采血量, 无法根据历史数据来科学预测采血量来满足临床需求, 同时避免偏型和过期浪费现象。

表 1 ABOAB 红细胞拟合模型及参数

分类	模型类型	平稳的 R 方	MAPE	MAE	MaxAE	正态化的 BIC	Sig
A 型	ARIAM(0,1,1)(0,1,1)	0.58	8.356	150.692	426.82	15.242	0.507
B 型	ARIAM(1,1,1)(1,1,1)	0.723	9.935	135.678	531.94	10.708	0.256
O 型	ARIAM(0,1,1)(0,1,1)	0.632	21.537	189.445	2024.2	11.089	0.804
AB 型	ARIAM(2,1,1)(1,1,1)	0.652	14.318	70.011	204.46	17.391	0.182
总血量	ARIAM(0,1,1)(0,1,1)	0.614	8.204	433.959	2364.9	12.931	0.471

表 2 4 种血型红细胞模型 1 至 6 月预测值与实际值U

指标	1 月	2 月	3 月	4 月	5 月	6 月
A-模型						
预测值	1 924	1 494	1 894	1 904	2 039	1 968
实际值	1 911	1 609	2 077	1 858	2 256	2 115
UCL	2 289	1 863	2 266	2 279	2 418	2 349
LCL	1 559	1 126	1 522	1 529	1 661	1 586
相对误差/%	0.07	7.14	8.81	2.47	9.61	6.95
B-模型						
预测值	1 596	1 350	1 661	1 628	1 696	1 673
实际值	1 493	1 413	1 722	1 597	1 771	1 582
UCL	1 923	1 677	1 998	1 970	2 044	2 027
LCL	1 269	1 023	1 325	1 286	1 348	1 320
相对误差/%	6.89	4.45	3.44	5.92	4.23	5.41
O-模型						
预测值	2 038	1 618	2 144	2 093	2 094	1 948
实际值	1 957	1 557	2 034	1 981	2 026	2 035
UCL	2 544	2 125	2 652	2 601	2 604	2 458
LCL	1 533	1 112	1 636	1 584	1 585	1 437
相对误差/%	4.13	3.91	5.40	5.65	3.24	4.27
AB-模型						
预测值	584	503	627	620	604	612
实际值	552	547	684	606	566	572
UCL	748	667	793	786	771	779
LCL	420	339	460	453	438	445
相对误差/%	5.79	8.04	8.33	2.31	6.71	6.99
总量-模型						
预测值	5 987	4 900	6 244	6 137	6 311	6 034
实际值	5 913	5 126	6 517	6 042	6 569	6 304
UCL	7 007	5 935	7 294	7 201	7 390	7 127
LCL	4 968	3 866	5 194	5 072	5 232	4 941
相对误差/%	1.25	4.40	4.18	1.57	4.65	4.28

时间序列方法是一种考虑对象本身的历史数据随时间发展变化的规律,并用该变量以往的资料建立统计模型来做外推的预测方法,在医疗卫生领域有广阔的应用前景^[3]。ARIMA 模型是一种预测精确度较高的模型。它是有时间变动的序列提出的建模方法,对每个周期中的同时间点的序列值进行分析,提取季节变化趋势,并对每个季节周期内部序列值提取非季节性的成分来做预测^[4]。各种条件只能在一定时期内保持稳定,因此预测时间不宜过长,应在动态变化中随时调整修正,以保持预测精度。同时,其缺点是计算相对复杂,因此,ARIMA 模型可以和其他时间序列分析方法如指数平滑法相互补充,配合使用。

由于我站对 A、B、O、AB 4 种血型 Rh 阴性红细胞在平常工作中在 6 d 未使用的新鲜血进行冰冻处理,保存期可长达 10 年,且库存量充足,所以不对 Rh 阴性红细胞进行预测研究。本研究运用 ARIMA 模型对我站 2008-01—2015-12 Rh 阳性悬浮红细胞月总供应量、A、B、O、AB 4 种血型的总共供应量进行预测。预测结果显示将预测值与实际值进行比较,实际值均落入预测值 95%的可信区间内,且平均误差相对较小,模型预测效果良好。因此,我们建立的模型可以用来中等血站作为供血需求量的预测,有利于及早采取针对性措施,增强采供血工作的主动性和预见性。但是 ARIMA 模型是依靠过去的统计资料建立的模型,并不能考虑到