

# 南宁市 2000—2009 年道路交通伤害时间序列分析\*

彭振仁<sup>1</sup> 杨莉<sup>1</sup> 刘勇<sup>2</sup> 张海英<sup>1</sup> 陈世艺<sup>1</sup> 尹晔<sup>1</sup> 覃莉<sup>2</sup>

**摘要:**目的 建立广西壮族自治区南宁市道路交通伤害的预测模型,掌握南宁市道路交通伤害的发生和变化趋势,为预防和控制南宁市道路交通伤害提供参考依据。方法 收集南宁市 2000—2009 年道路交通伤害资料,进行时间序列分析,建立自回归求和移动平均模型(ARIMA 模型),对南宁市 2010—2011 年道路交通伤害发生情况进行预测。结果 建立了南宁市道路交通伤害事故发生次数、死亡人数、受伤人数和直接经济损失各自的 ARIMA 模型,模型拟合与预测效果良好,预测模型均为 ARIMA(1 0 0)。2010 年各指标的预测值依次为 472 次、145 人、562 人、157.043 6 万元;2011 年各指标的预测值依次为 464 次、141 人、527 人、161.120 9 万元。结论 ARIMA 模型在道路交通伤害预测中具有较好的应用价值。

**关键词:** 道路交通伤害; 时间序列分析; 自回归求和移动平均模型

中图分类号: R 181.3<sup>+</sup>9

文献标志码: A 文章编号: 1001-0580(2011)02-0574-02

**Time series analysis on road traffic injury in Nanning city 2000 - 2009** PENG Zhen-ren, YANG LI, LIU Yong, et al. *Department of Occupational and Environmental Health, College of Public Health, Guangxi Medical University (Nanning 530021, China)*

**Abstract: Objective** To establish predictive models for road traffic injury (RTI) in Nanning city and to elucidate the trend of RTI for prevention and control of RTI in Nanning. **Methods** The RTI data from 2000 to 2009 in Nanning city were collected and the autoregressive integrated moving average (ARIMA) model was used to analyze and predict the trend of RTI from 2010 to 2011 in Nanning. **Results** A series of predictive equations on RTI were established based on ARIMA model. The model fitting was effective and the predictive data on RTI were close to the true statistical data and all the predictive models were ARIMA(1 0 0). The predictions for number of accident, number of people dead and injured due to injury and direct economic lost were 472, 145 and 562 persons and 1 570 436 RMB yuan for the calendar year of 2010 and 464, 141 and 527 persons and 1 611 029 RMB yuan for 2011 based on the models. **Conclusion** The ARIMA model could be applied in RTI prediction effectively.

**Key words:** road traffic injury (RTI); time series analysis; autoregressive integrated moving average model

道路交通伤害(road traffic injury, RTI)已成为全球重大公共卫生问题,每年超过 100 万人死亡<sup>[1]</sup>,伤害的数目高达 5 000 万<sup>[2]</sup>,直接经济损失为 5 180 亿美元<sup>[3]</sup>。据国家公安部交通管理局统计,从 2001—2007 年中国每年发生道路交通事故约 50 万起,因道路交通事故死亡人数约 10 万人,占全世界的 16% 左右,居世界第一。随着机动车数量的持续增加,RTI 已成为广西突出的公共卫生问题,给社会和家庭带来巨大损失。有研究表明,广西城市运输事故导致的死亡位于居民各类伤害死亡之首<sup>[4]</sup>。本研究通过收集南宁市 2000—2009 年 RTI 资料,进行时间序列分析,建立自回归求和移动平均模型 (autoregressive integrated moving average model, ARIMA 模型),对南宁市 2010—2011 年道路交通伤害发生情况进行预测,为道路交通伤害的预防、控制、监测以及制定干预措施和决策立法提供科学依据。

## 1 资料与方法

**1.1 资料来源** 资料来源于广西公安厅交警总队记录在案南宁市 2000—2009 年 RTI 资料,RTI 资料主要通过交通事故信息采集表、统计表、月报表和死亡事故情况分析表等来收集。RTI 的车祸判断、分类标准、车祸死亡判断均按公安部统

一规定。

**1.2 方法** 采用时间序列分析预测 ARIMA 模型<sup>[5]</sup>。该模型有 3 种基本模式:自回归模型 [AR(p)]、移动平均模型 [MA(q)] 与自回归移动平均模型 [ARMA(p, q)]。当序列为平稳序列,其模型表达式为:  $y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$ ,  $p, q$  为自回归和移动平均的阶数,相应参数  $\varphi_1, \varphi_2, \dots, \varphi_p$  与  $\theta_1, \theta_2, \dots, \theta_q$  分别为自回归和移动平均系数。

运用 ARIMA 方法的前提条件是:作为预测对象的时间序列是一零均值的平稳随机序列,平稳随机序列的统计特性不随时间的推移而变化<sup>[6]</sup>。当序列为非平稳序列,但通过  $d$  次差分可使序列平稳时,采用的模型称作自回归求和移动平均模型,即 ARIMA(p, d, q) 模型,鉴于 3 种基本模式可视作 ARIMA 模型的特例,故 ARIMA 模型又常被作为这一族模型的总称<sup>[5]</sup>。本研究的时间序列分析指标为:每年 RTI 的发生次数、死亡人数、受伤人数和直接经济损失。

**1.3 统计分析** 应用 PASW Statistics 18.0 统计软件对南宁市 2000—2009 年 RTI 资料进行时间序列分析,分析流程为:建立时间序列→模型识别→参数估计→模型诊断→模型预测。

## 2 结果

**2.1 建立时间序列(表 1)** 2000—2005 年交通事故次数、死亡人数、受伤人数、直接经济损失各指标总体呈上升趋势,2006—2009 年总体呈下降趋势。

**2.2 模型识别** 将 2000—2009 年交通事故次数、死亡人数、受伤人数、直接经济损失各指标分别作时间序列图,可知各指标不随时间的变化而变化。对 RTI 各指标的时间序列作自相

\* 基金项目:国家自然科学基金(30860237);广西自然科学基金(桂科自 0832156);广西 2009 年研究生教育创新计划资助项目(2009105981004M172)

作者单位:1. 广西医科大学公共卫生学院职业卫生与环境健康教研室 广西 南宁 530021; 2. 广西医科大学研究生院  
作者简介:彭振仁(1982—),男,广西桂林人,硕士在读,研究方向:伤害的预防与控制。

通讯作者:杨莉, E-mail: yangli8290@hotmail.com

关图( autocorrelation funtion ,ACF) 和偏自相关图( partial auto-correlation function ,PACF) ,各指标的时间序列所对应的 ACF 和 PACF 均拖尾 ,自相关系数和偏自相关系数在滞后 8 阶后仍落在置信区间内 ,可判定各指标的时间序列均为平稳随机序列。从 ACF 和 PACF 可看出 ,各指标的时间序列自相关系数和偏自相关系数基本是在滞后 1 阶后迅速减小 ,可初步估计各指标的时间序列的 ARIMA 模型的自回归阶数  $p \leq 1$  和移动平均阶数  $q \leq 1$ 。综合上述分析过程 ,可初步判定各指标的时间序列的 ARIMA 模型为 ARIMA( 1  $\rho$  0) 、ARIMA( 0  $\rho$  1) 或 ARIMA( 1  $\rho$  1)。

表 1 2000—2009 年南宁市 RTI 发生情况

年 份	交通事故次数(次)	死亡人数(人)	受伤人数(人)	直接经济损失(万元)
2000	208	78	177	70.997 0
2001	271	103	225	88.434 4
2002	229	126	199	85.011 0
2003	246	132	212	92.112 8
2004	1 038	165	904	193.903 2
2005	1 532	267	1 355	403.793 6
2006	1 142	239	1 458	206.121 3
2007	749	203	940	180.123 6
2008	691	170	837	259.394 0
2009	483	150	603	145.745 1

2.3 模型参数估计(表 2) 对交通事故次数、死亡人数、受伤人数、直接经济损失各指标的时间序列的 ARIMA( 1  $\rho$  0) 、ARIMA( 0  $\rho$  1) 和 ARIMA( 1  $\rho$  1) 模型进行参数的估计及假设检验 ,发现只有当 ARIMA( 1  $\rho$  0) 模型对应的平均绝对百分比误差( MAPE) = 31.988、最大绝对百分比误差( MaxAPE) = 89.545 和标准化贝叶斯信息准则( Normalized BIC) = 10.561 最小 ,且各指标的自回归系数估计值差异均有统计学意义(  $P < 0.05$  ) ,故 ARIMA( 1  $\rho$  0) 模型是最优选择。

表 2 ARIMA( 1  $\rho$  0) 模型参数估计

模 型	自回归系数估计值	t 值	P 值
交通事故次数	0.727	2.867	0.021
死亡人数	0.928	6.011	0.000
受伤人数	0.865	4.520	0.002
直接经济损失	0.361	2.086	0.039

2.4 模型诊断 对交通事故次数、死亡人数、受伤人数、直接经济损失各指标的时间序列的残差作序列图、ACF 和 PACF。残差序列图显示残差序列随机分布 ,残差序列的 ACF 和 PACF 均拖尾 ,残差序列自相关系数和偏自相关系数在滞后 8 阶后仍落在置信区间内 ,且残差序列统计量( Box-Ljung statistic) 均无统计学意义(  $P > 0.05$  ) ,提示 ARIMA( 1  $\rho$  0) 模型拟合较好 ,可以用此模型来进行预测。

2.5 模型预测 根据以上结果可获得交通事故次数、死亡人数、受伤人数、直接经济损失各自 ARIMA( 1  $\rho$  0) 模型的预测公式:  $Y_{(t)}_{\text{交通事故次数}} = 441.912 + 0.727Y_{(t-1)}_{\text{交通事故次数}}$ ;  $Y_{(t)}_{\text{死亡人数}} = 86.978 + 0.928Y_{(t-1)}_{\text{死亡人数}}$ ;  $Y_{(t)}_{\text{受伤人数}} = 300.825 + 0.865Y_{(t-1)}_{\text{受伤人数}}$ ;

$Y_{(t)}_{\text{直接经济损失}} = 163.423 + 0.361Y_{(t-1)}_{\text{直接经济损失}}$ 。根据预测公式可对南宁市 2010—2011 年 RTI 的交通事故次数、死亡人数、受伤人数、直接经济损失 4 项指标进行预测。2010 年各指标预测值( 95% 可信区间) 依次为 472( - 387 ~ 1 331) 次 ,145( 43 ~ 248) 人 ,562( - 245 ~ 1 370) 人 ,157.043 6( - 78.268 0 ~ 392.355 3) 万元; 2011 年各指标预测值依次为 464( - 599 ~ 1 526) 次 ,141( 1 ~ 281) 人 ,527( - 541 ~ 1 595) 人 ,161.120 9( - 89.043 5 ~ 411.285 2) 万元。

3 讨 论

RTI 的发生受人、车、环境、道路等多种因素的影响 ,要获得理想的预测模型比较困难 ,常用的统计预测有定性预测、一元线性回归方法、多元回归方法、非线性方法、趋势外推法、分解分析方法、指数平滑法、灰色预测法、ARIMA 预测方法、景气预测方法、状态空间模型等。时间序列分析法是根据过去的变化趋势预测未来的发展 ,其前提是假定事物的过去延续到未来。时间序列预测法可用于短期预测、中期预测和长期预测 ,一般情况下 ,时间序列分析法对于短、近期预测效果较好 ,但如延伸到更远的将来 ,则存在很大的局限性 ,导致预测值偏离实际较大而使决策失误。因本研究 RTI 资料时间序列长度为 10 ,只对未来 2 年的 RTI 情况进行预测 ,而时间序列 ARIMA 预测方法是在短时间内不包括季节变动的反复预测 ,故采用 ARIMA 模型对南宁市 2000—2009 年 RTI 的交通事故次数、死亡人数、受伤人数和直接经济损失作时间序列分析 ,并对 2010—2011 年的交通事故次数、死亡人数、受伤人数和直接经济损失进行预测分析。建立 ARIMA(  $p$   $d$   $q$  ) 模型的难点是确定合适的  $p$ 、 $d$ 、 $q$  值 ,使模型参数具有统计学意义 ,模型最优化。而模型的拟合过程亦很复杂和繁琐 ,因此 ,要构建正确的预测模型必须严格遵循建模原则 ,步骤明确 ,反复探索。本研究经过反复拟合模型 ,最终得出最优模型 ARIMA( 1  $\rho$  0) ,此模型等价于 AR( 1) 的自回归过程 ,服从 AR( 1) 的线性模型 ,序列中的当前观测值和它的上一个观测值相关 ,但与更前面的其他观测值均无关。本次预测结果显示预测值均落在 95% 可信区间内 ,预测值与 2009 年的实际值较接近 ,预测值未出现大幅波动 ,说明模型的建立成功。本研究时间序列分析未分离出 RTI 的季节变化因素 ,说明 RTI 不随季节变化而波动。

参考文献

- (1) Mock C ,Kobusingye O ,Anh LV ,et al. Human resources for the control of road traffic injury [J]. Bulletin of World Health Organization , 2005 83( 4) : 294.
- (2) 世界卫生组织. 世界预防道路交通伤害报告 [M]. 刘光远 ,译. 北京: 人民卫生出版社 2004: 3.
- (3) Tacobs G ,Aeron-Thomas A ,Astrop A. Estimating global road fatalities [M]. Crowthorne: Transport Research Laboratory 2000: 445.
- (4) Yang L ,Lam LT ,Liu Y ,et al. Epidemiological profile of mortality due to injuries in three cities in the Guangxi province ,China [J]. Accident Analysis and Prevention 2005 37( 1) : 137 - 141.
- (5) 方积乾 ,陆盈. 现代医学统计学 [M]. 北京: 人民卫生出版社 , 2002: 233.
- (6) 孙振球 ,徐勇勇. 医学统计学. 2 版. [M]. 北京: 人民卫生出版社 2008: 463.

收稿日期: 2011-12-21

( 张翠编校)