2021 年 3 卷 10 期 ISSN:2705-0440(Print); 2705-0459 (Online)

1996 年 ~2019 年江苏省恶性肿瘤死亡率预测研究

张蓓蓓
南京医科大学康达学院，江苏 连云港 222000

摘　要：目的：根据江苏省恶性肿瘤死亡变化趋势，结合其死亡率统计资料，建立 ARIMA 模型并对江苏省恶性肿瘤死亡率进行预测。方法：以 SPSS25.0 软件为工具，对 1996 年至 2015 年江苏省年度死亡率数据建模，预测 2016 年至 2019 年恶性肿瘤死亡率，并与实际数据对比，进行预测效果分析。结果：ARIMA(1,2,0) 模型的拟合系数 R2=0.848，与实际统计数据的拟合水平较高，且残差序列满足白噪声检验要求，预测结果的均方根误差 MAPE 为 2.05%。结论：ARIMA 模型拟合及预测效果良好，能够较好地描述该时段江苏省恶性肿瘤死亡情况，为制定恶性肿瘤疾病防控方案，落实防控措施提供较为准确的数据支持。

关键词：恶性肿瘤；ARIMA 模型；时间序列；预测

恶性肿瘤是肌体在致瘤因素的作用下，因遗传物质病变而引起的组织细胞异化或过度增生所形成的不良新生物。恶性肿瘤不仅局部快速增生，同时能够感染正常肌体组织，进而造成在人体内扩散，对肌体产生严重的危害 [1]。根据国家癌症中心发布《2019 年全国癌症报告》，近年来恶性肿瘤的发病死亡均呈整体上升态势，我国恶性肿瘤死亡占居民全部死因的 23.91%，恶性肿瘤已经成为严重威胁中国人民生命健康的公共卫生问题之一。

江苏省作为恶性肿瘤高发省份之一，《2019 年江苏省卫生健康事业发展统计公报》中的数据显示，江苏省恶性肿瘤死亡率为 206.85/10 万，占死亡构成的 30.2%，成为江苏省居民死因中的首位。开展恶性肿瘤死亡率预测研究是分配卫生资源和开展恶性肿瘤防治工作提供重要依据。死亡率预测研究多建立在统计数据的基础上，基于此，江苏省对恶性肿瘤发病、死亡情况的数据统计工作尤为重视，参照国家肿瘤登记年报数据纳入的原则和标准，逐步完善了恶性肿瘤的数据统计和筛选机制，积累的相关统计资料为本文研究提供了数据支撑 [2,3]。本文运用求和自回归移动平均 (autoregressive integrated moving average，ARIMA) 模型对江苏省恶性肿瘤年度死亡率进行拟合和预测，选取合理数据评价指标对模型预测效果进行分析，为制定恶性肿瘤防治策略提供有效的参考。

1　资料与方法

1.1 资料来源

根据江苏省卫生健康委员会编写的 1996 年 ~2019 年《江苏卫生计生统计年鉴》中恶性肿瘤年度死亡率统计数据进行死亡率预测模型的筛选及检验，其中 1996 年 ~2015 年恶性肿瘤死亡率数据用于拟合死亡率的 ARIMA 时间序列模型，2016 年 ~2019 年恶性肿瘤死亡率数据作为检验模型预测效果的对比值。

1.2 建模方法

ARIMA 模型是由 Box 和 Jenkins 于 70 年代初提出的著名短期时间序列预测方法，又称为 Box-Jenkins 模型，是一种适用于平稳性时间序列的预测模型 [4,5]。该模型的一般表达式为 ARIMA(p,d,q)，其中参数 p、q 和 d 分别表示序列的自相关函数 (ACF)、偏自相关函数 (PACF) 的阶和进行差分的次数。

本文在江苏省历年恶性肿瘤死亡率数据汇总的基础上，通过统计分析软件 SPSS 25.0 对进行数据分析、建模、评估和短期预测。

ARIMA 建模过程分为以下四个步骤 [4]：(1) 序列预处理：对非平稳的初始时间序列需先进行数据预处理，通过一定的变换与差分，转化为符合要求的平稳时间序列；(2) 模型识别与参数确定：结合预处理后的平稳序列的自相关函数 (ACF) 和偏自相关函数 (PACF) 图，估计 ARIMA 模型中的 p、q 的值；(3) 参数估计及判定：运用最大似然估计法估计模型的系数并进行检验。模型的残差序列需满足白噪声检验 [6]，结合拟合系数 R2、BIC 最小信息准侧等，筛选出拟合效果最好的模型；(4) 模型的预测：利用建立的 ATIMA 模型进行数据预测，以误差分析指标评估模型的预测效果。

2　结果

2.1 死亡趋势分析及序列平稳化

图 1 是江苏省 1996 年 ~2019 年的恶性肿瘤死亡率的时间序列图，图中显示， 1996 年 ~2000 年恶性肿瘤的死亡率呈现小幅下降，在 2000 年以后恶性肿瘤的死亡率呈现整体上升趋势。由此可知，该时间序列既有长期上升的趋势性，又有短期变动的波动性，为非平稳序列，需要进行数据的预处理。对原始序列进行二阶差分（d=2）后，如图 2 所示，预处理后的序列在 0 附近呈现平稳的小幅上下波动，表现为平稳性的典型特征 [6]。

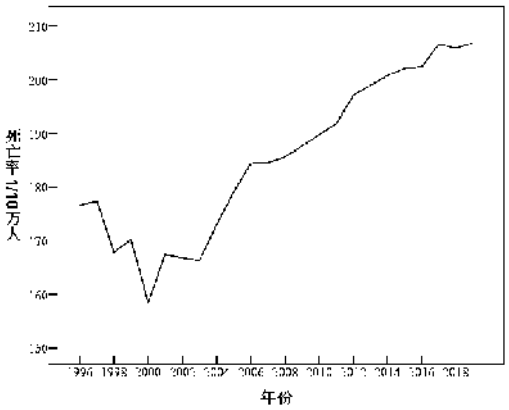

图 1 江苏省恶性肿瘤死亡率时序图

图 2 二阶差分后的时间序列图

2.2 模型的识别和定阶

经过二阶差分后数据平稳，可确定模型的参数 d=2。平稳序列的 ACF（图 3）和 PACF 分析（图 4），延迟二阶、一阶后，收敛于置信区间内，因此 p、q 分别在 0、1，0、1、2 中进行选取。由 2 个参数的不同选择可得到六种备选模型，对六个模型分别进行参数估计和残差白噪声诊断，结合拟合系数 R2、最小信息准则 BIC 等原则筛选出最优模型。

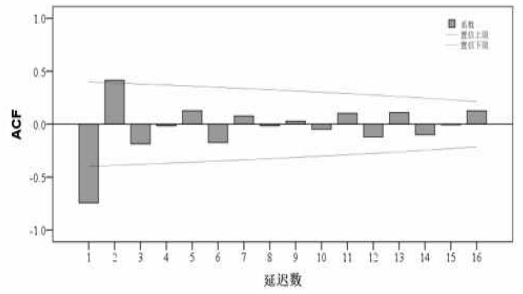

图 3 平稳序列自相关 (ACF) 图

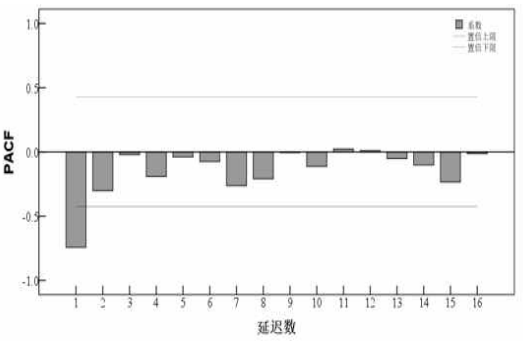

图 4 平稳序列偏自相关 (PACF) 图

2.3 参数估计及诊断

采用 SPSS 25.0 软件中时间序列预测分析模块，对拟定备选模型分别进行拟合分析。基于模型系数的 T 检验要求，剔除不满足要求的模型，得到模型 ARIMA(1,2,0) 结果详见表 1。由表 1 可知，模型系数通过 t 检验 (P<0.05)；在拟合效果方面，模型的拟合系数 R2=0.848，模型的拟合程度较高。

对模型 ARIMA(1,2,0) 的残差序列进行白噪声检验，经统计，残差序列在各滞后阶数下 Q 统计量的 P 值均大于 0.05，故模型的残差序列为白噪声序列 [7]。由此可知，模型 ARIMA(1,2,0) 能够满足检验要求，可进一步分析模型的预测效果以确定其适用性。

表 1 ARIMA 模型参数检验及拟合结果表

变量	ARIMA(1,2,0)		
	系数	t 值	P 值
AR(1)	-0.799	-5.613	<0.001
常数	216.737	0.772	0.452
Sig			
R2		0.848	
BIC		3.944	

模型 ARIMA(1,2,0) 的拟合数据与实测数据曲线对比图详见图 5，由图中可见，模型不仅能拟合原始数据的整体发展趋势，同时能够反映数据的局部小幅变化的波动性，表现出较好的拟合效果。

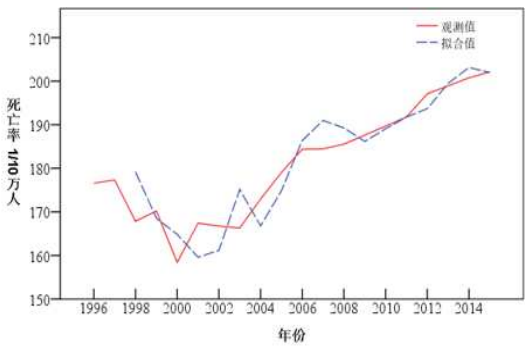

图 5 ARMA(1,2,0) 模型拟合效果图

2.4 误差指标选取及预测效果分析

本文选取绝对误差、误差率等误差对比指标对模型的预测效果进行分析，采用绝对误差 (E)、绝对误差率 (ER) 和均方根误差 (MAPE) 进行综合评价。各误差指标的计算结果越小，模型的预测精度也就越高，且一般认为 MAPE < 10% 的情况下，预测精度较高。

ARIMA(1,2,0) 模型对 2016 年 ~2019 年短期预测数据相应误差计算指标汇总详见表 2。由表 2 中数据可知，模型的均方根误差 MAPE 为 2.05%，表明模型的整体预测效果较为理想；从每年的绝对误差率来看，最大值为 2019 年的 3.78%，预测精度较高，满足对恶性肿瘤死亡率发展水平情况的整体评估，对具体防控措施的制定和医疗物质的分配均可提供较

为准确的数据支持。

表 2 ARIMA 模型预测结果比较

时间	实际死亡率 1/10 万	预测死亡率 1/10 万	E	ER	MAPE
2016 年	202.37	202.64	0.27	0.13%	2.05%
2017 年	206.63	202.42	4.21	2.04%	
2018 年	205.84	201.22	4.62	2.24%	
2019 年	206.85	199.03	7.82	3.78%	

3 讨论

根据国家癌症中心 2019 年发布的我国恶性肿瘤流行情况分析报告，在 2015 年我国恶性肿瘤总体死亡人数约为 233.8 万人，每年为医治恶性肿瘤所需的医疗花费超过 2200 亿元。进十年来，我国恶性肿瘤死亡率基本上每年保持 2.5% 左右的涨幅。由本文对于江苏省恶性肿瘤死亡率的预测分析可知，江苏省恶性肿瘤死亡率总体保持上升趋势。近几年，得益于整体医疗水平的提高和居民对恶性肿瘤危害性认识的逐渐加强，恶性肿瘤的死亡率增幅较小，近十年江苏省恶性肿瘤死亡率的年度增长平均为 1.13%，低于全国水平。从总体死亡率水平上来看，2014 年 ~2019 年江苏省恶性肿瘤的死亡率不低于 200/10 万人，依然在江苏省居民死因中占据首位，造成总体医疗负担的持续加重。因此，开展恶性肿瘤死亡率的统计研究，实现医疗资源的科学配比，是恶性肿瘤防治工作中必不可少的参考。

关于死亡预测模型研究，回归分析模型和时间序列分析模型在适用性和准确性方面效果最好 [8]。从恶性肿瘤死亡率发展规律来看，时间序列是包含整体趋势性和局部波动性的非平稳时间序列。ARIMA 模型能剔除时间序列的趋势性影响，已经广泛运用在各类疾病的发病及死亡情况的预测研究中。本文以 1996 年 ~2015 年江苏省恶性肿瘤发病数据作为拟合样本得到了 ARIMA(1,2,0) 模型，从拟合结果来看，模型的整体拟合效果良好，对死亡率发展趋势的描述与实测数据基本一致，能够较好的反映江苏省恶性肿瘤发病的整体规律。由预测结果的对比分析可知，模型对死亡率预测整体精度高，能满足恶性肿瘤防治工作的要求，可用于我国恶性肿瘤死亡率情况的短期预测。

参考文献

[1] 娄长丽 . 临床常见恶性肿瘤［M］. 吉林 : 吉林科学技术出版社，2009：6 － 10.

[2] 国家癌症中心 . 中国肿瘤登记工作指导手册 (2016)[M]. 北京：人民卫生出版社，2016，59-75.

[3] 韩仁强，武鸣，缪伟刚，等. 2015 年江苏省恶性肿瘤发病和死亡分析 [J]. 中国肿瘤，2020，29(2) : 81 － 89.

[4] 丁勇，张蓓蓓，吴静，等 . ARIMA 乘积季节模型预测我国戊肝的发病趋势 [J]. 南京医科大学学报 (自然科学版), 2020,40(11):1725-1729.

[5] 于林凤，吴静，周锁兰，丁勇 . ARIMA 季节模型在我国丙肝发病预测中的应用 [J]. 郑州大学学报 (医学版), 2014，49(03)：344-348.

[6] 庞艳蕾，张惠兰，李向云，等 . 灰色模型 GM(1,1) 和 ARIMA 在拟合全国婴儿、5 岁以下儿童死亡率中的应用 [J]. 中国卫生统计，2015,32(03)：461-463.

[7] 刘洁 , 高茵茵 , 曲波 , 等 . 应用 ARIMA 模型预测我国孕产妇死亡率 [J]. 中国医科大学学报，2011,40(02):107-108+121.

[8] 黄国宝，黎衍云，吴菲 . ARIMA 模型和 ARIMA-SVM 模型对上海市 2 型糖尿病患者肺结核发病的预测效果 [J]. 复旦学报 (医学版),2020,47(06):899-905.

[基金项目]：2019年度江苏省高等学校自然科学研究面上项目（江苏省恶性肿瘤流行特征分析与死亡预测研究19KJD330001）