

指数平滑法在麻疹发病率预测中的应用

唐广心¹, 张飞飞², 鲁苇葭³, 彭敬¹, 贺利敏¹, 李苑¹

1. 深圳市宝安区疾病预防控制中心, 广东 深圳 518101; 2. 深圳市宝安区松岗预防保健所; 3. 四川大学公共卫生学院

摘要: **目的** 通过指数平滑法对麻疹发病率进行预测, 为制订预防和控制麻疹的流行提供依据。 **方法** 收集 2005–2015 年某区确诊的麻疹发病率, 通过 SPSS 18.0 进行指数平滑法预测, 并对预测结果进行评价。 **结果** 指数平滑法的 Holt–Winters 相乘模型适合麻疹月发病率的预测 ($RMSE=7.69$, $\text{adj } R^2=0.52$, $R^2=0.85$, Ljung–Box $Q=21.91$), 预测模型线性趋势 $\text{Alpha}=1.000$, $P=0.000$ 。模型预测结果显示 2015 年的麻疹疫情处于相对低发水平, 麻疹月发病率将呈上升后波动趋势, 实际麻疹月发病率则出现先上升后急剧下降维持在低水平上的波动趋势。 **结论** 指数平滑模型 1~5 月份预测结果与实际发病率的相对误差仅为 10.60%, 说明指数平滑法在短期预测麻疹发病情况中有一定效果。

关键词: 指数平滑法; 麻疹; 预测

中图分类号: R511.1 **文献标识码:** B **文章编号:** 1006–3110(2018)06–0757–03 **DOI:** 10.3969/j.issn.1006–3110.2018.06.034

Application of exponential smoothing method to forecasting the incidence rate of measles

TANG Guang-xin*, ZHANG Fei-fei, LU Wei-jia, PENG Jing, HE Li-min, LI Yuan

* Baoan District Center for Disease Control and Prevention, Shenzhen, Guangdong 518101, China

Corresponding author: LI Yuan, E-mail: 395551194@qq.com

Abstract: **Objective** To predict the incidence rate of measles by using exponential smoothing method, and to provide a basis for developing measles prevention and control strategies. **Methods** The data about the incidence rate of confirmed measles cases in a district from 2005 to 2015 were collected. Exponential smoothing prediction was performed by using SPSS18.0 statistical software, and then the prediction results were evaluated. **Results** Holt–Winters multiplicative model was suitable for forecasting the monthly incidence rate of measles ($RMSE=7.69$, $\text{adj } R^2=0.52$, $R^2=0.85$, Ljung–Box $Q=21.91$). As for Holt linear trend model, Alpha and P values were 1.000 and 0.000 respectively. Forecasts by the model showed that the level of measles incidence in 2015 was relatively low, and the monthly incidence rates of measles exhibited fluctuations after going through an increase trend. The monthly actual incidence rates of measles displayed low-level fluctuations after undergoing an increase trend and a steep decline trend. **Conclusions** Forecasts by exponential smoothing model reveal that the relative error for the monthly predicted and actual incidence rates in January–May is only 10.60%, indicating that exponential smoothing method is adapted to predicting the short-term incidence of measles.

Key words: exponential smoothing method; measles; forecasting

继天花和脊灰消灭后, WHO 在世界范围内提出了消除麻疹分步实施的规划, 目前 WHO 美洲区已于 2006 年成功消除麻疹。2006 年卫生部出台了《2006–2012 年全国消除麻疹行动计划》(以下简称《行动计划》), 中国消除麻疹活动正式启动^[1]。中国向 WHO 承诺于 2012 年将麻疹发病率控制在百万分之一以下, 时至今日, 麻疹疫情仍然在我国肆虐, 特别是某区麻疹发病一直维持在百万分之六十左右, 距离消除目标仍有较大差距。为了解某区麻疹发病规律及变化趋势,

本次研究采用 2005–2015 年某区麻疹发病资料, 采用指数平滑法(exponential smoothing, ES)拟合该区 2005–2014 年麻疹月发病率的变化趋势, 建立该区麻疹发病率预测模型, 希望能为 2015 年麻疹疫情的预警提供客观信息。

1 资料与方法

1.1 资料来源 数据来源为 2005 年 1 月–2015 年 12 月“中国疾病预防控制中心信息系统”的某区麻疹月报表, 人口数资料来源于该区公安局。

1.2 研究方法 指数平滑法是布朗(Robert G. Brown)所提出, 是用序列以往实际观测值的加权平均来预测未来数值, 序列中近期的数据赋予较大的权

基金项目: 宝安区科技局立项课题(2013091, 2016CX226)

作者简介: 唐广心(1978–), 男, 硕士, 副主任医师, 主要从事流行病与计划免疫工作。

通信作者: 李苑, E-mail: 395551194@qq.com。

重,远期的数据赋予较小的权重,由于随着时间的流逝,过去值的影响逐渐减小^[2-3],主要用于中短期预测^[4-5]。指数平滑法在模型拟合时采用了简单季节、Holt-Winters 相加、Holt-Winters 相乘三种模型,每种模型均进行了参数的统计检验,给出了多个拟合优度指标。在参数选择过程中,使 RMSE(均方根误差)最小的模型为最优模型。平稳的 R^2 统计量用于比较模型中的固定成分和简单均值模型的差别,取正值时标示模型优于简单均值模型,取负值时表示劣于简单均值模型,为零时表示和简单均值模型效果相同。 R^2 表示模型所能解释的数据变异占总变异的比列,该数据越大表示结果越佳。正态化的 BIC 表示的是模型对数据的释度,该统计量基于均方误差统计量,并考虑了模型的参数个数和序列数据个数,BIC 值越小相对模型越优。Ljung-Box Q 是检验残差序列是否为独立序列的假设检验,当假设检验的 $P>0.05$ 时,说明数据拟合后的残差序列不存在自相关,可以采用指数平滑序列模型。在模型选择过程中,根据均方根误差 (root-mean-square error, RMSE) 和平均绝对误差绝对值 (mean absolute percent error, MAPE) 最小原则确立最优模型^[6]。统计分析利用 PASW statistics 18.0 对数据进行定义、拟合和分析。

2 结果

2.1 指数平滑法模型的选择 根据某区 2005-2015 年的数据发现,每年的 5-9 月麻疹发病水平较高,说明麻疹发病具有一定的季节性。其中 2007 年麻疹月发病率明显高于其他年份。自 2009 年起,麻疹发病一直维持在较低的水平,但在 2012 年后出现了逐年上升的趋势。利用上述数据通过指数平滑模型拟合与对 2015 年趋势的预测。根据三种拟合方法的参数检验结果及上述原则,采取 Holt-Winters 相乘模型拟合数据,拟合结果平稳的 $R^2 = 0.52$, $R^2 = 0.85$, RMSE = 7.69,由于因变量数据为季节性数据,因此平稳的 R^2 更具有代表性。从两个 R^2 值及 RMSE 值来看,该指数平滑模型的拟合情况比较良好,见表 1。

表 1 指数平滑法三种的拟合结果分析表

模型类型	平稳的 R^2	R^2	RMSE	正态化的 BIC	Ljung-Box 统计量	DF	Sig.
简单模型	0.45	0.78	9.34	4.55	27.89	15	0.03
Holt-winters 相加模型	0.51	0.84	7.99	4.28	34.52	15	0.00
Holt-winters 相乘模型	0.52	0.85	7.69	4.20	21.91	15	0.11

2.2 采用 Holt-Winters 相乘模型拟合效果图及参数

检验 模型将拟合结果和实际数据进行比较,同时将预测结果予以显示,见图 1。

表 2 给出了指数平滑法模型参数估计值列表,从表中可以看到本实验拟合 Holt-Winters 相乘模型的指数平滑模型的水平 $\text{Alpha}=1.000$, $P=0.000$,具有显著性。而季节 $\text{Delta}=1.000$,没有通过统计学检验,因此可以判断麻疹发病虽然具有季节性,但从长期看麻疹月发病率序列尚没有季节性特征。见表 2。

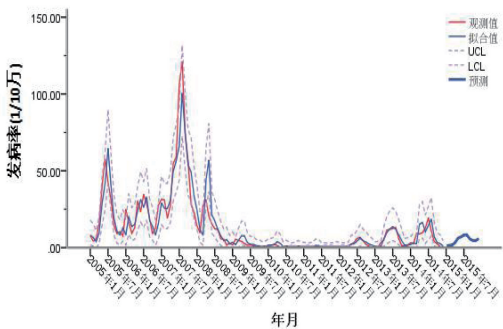


图 1 模型拟合结果与预测结果比较图

表 2 模型统计量分析表

模型类型	指标	估计	SE	t 值	P 值
Holt-Winters 相乘模型	Alpha(水平)	1.000	0.088	11.317	0.000
	Gamma(趋势)	0.000	0.016	0.002	0.998
	Delta(季节)	1.000	108.992	0.009	0.993

2.3 指数平滑 Holt-Winters 相乘模型的应用 模型预测结果提示 2015 年的麻疹月发病率处于历史较低水平,将呈现先上升后波动的状态,5-9 月份仍为全年的高发区,8 月份为最高峰,月发病率达 8.37/10 万,全年麻疹病例发病率 57.81/10 万。2015 年实际麻疹报告发病率 1-4 月份呈现出逐步升高、5 月份突然暴发,较 4 月份有 2 倍以上的增长,6 月份又有所下降,7 月份之后骤跌直到结束。2015 年的实际报告发病率与预测结果进行比较,可见模型预测结果与实际结果在 1-5 月份拟合较好,平均相对误差仅为 10.60%,而 6 月份之后,实际结果与模型预测结果差距甚远,其中原因将在讨论中进行深度阐述。结果见表 3。

表 3 用 Holt-Winters 相乘模型对宝安区 2015 年麻疹数据预测结果

月份	实际发病率 (1/10 万)	预测发病率 (1/10 万)	相对误差 (%)	预测发病率 95%可信 区间上限(1/10 万)	预测发病率 95%可信 区间下限(1/10 万)
1	1.11	1.09	-2.24	4.88	0.56
2	0.56	1.39	59.91	6.84	1.85
3	1.30	1.66	21.67	8.18	3.17
4	2.97	3.17	6.25	15.47	6.93
5	6.32	6.25	-1.04	30.27	14.71

续表 3

月份	实际发病率 (1/10 万)	预测发病率 (1/10 万)	相对误差 (%)	预测发病率 95% 可信 区间上限(1/10 万)	预测发病率 95% 可信 区间下限(1/10 万)
6	4.27	6.97	38.71	33.46	17.43
7	1.86	8.31	77.65	39.52	21.86
8	1.11	8.37	86.69	39.41	23.05
9	0.00	5.65	100.00	26.29	16.34
10	0.19	4.54	95.91	20.81	13.76
11	0.37	4.48	91.71	20.22	14.12
12	0.19	5.93	96.87	26.39	19.27
合计	20.25	57.81	672.07	271.74	153.05

3 讨 论

消除麻疹是我国当前重要的公共卫生任务,对其发病情况的预测研究具有深远意义,也是流行病学研究的重要课题之一。指数平滑法在预测中的研究应用历时已久,在理论有成熟的经验方法和明确的判别准则,适用于各种疾病的发病预测^[7-8]。指数平滑法作为时间序列分析的一种重要方法^[9-10],具有方法简洁、使用范围广泛、充分利用历史资料信息等优点。目前指数平滑过程可借助统计软件进行计算机操作完成,是一种简单易行预测方法,其实用性强、预测精确度较高,对于监测疾病的流行趋势,评价防控措施效果有一定的作用。如果实际发病率在预测值 95% 可信区间内,说明当月疫情基本正常;如果超出预测值 95% 可信区间,则说明当月疫情已不同于以往流行规律,应警惕传染病暴发或流行的可能^[11]。本次研究中 2015 年的麻疹发病率实际值绝大部分落在预测值的 95% 可信区间下限内,说明 2015 年的麻疹疫情处于相对低发水平。

本次研究分别拟合了简单季节、Holt-Winters 相加、Holt-Winters 相乘三种模型,根据指数平滑法的 RMSE(均方根误差)最小的原则,发现本次研究的月发病率数据采用 Holt-Winters 相乘模型拟合效果最好,与朱奕奕等^[12]用指数平滑法研究甲肝疫情时的结果一致。

本次研究建立的指数平滑法模型用于预测相对误差达 672.07%,从总体上看本次研究预测结果是失败的。分析各月的预测结果发现,1-5 月份预测结果的相对误差为 10.60%,此时预测结果良好,预测结果略高于实际发病率。在这个阶段,2、3 月份的预测结果与实际发病率差别较大,主要原因是 2、3 月份是我国传统的春节,该区作为一个流动人口聚集的大区,大批

的务工人员返乡,也造成发病人数减少。同时此阶段也处于麻疹疫情的低发期,时间序列预测模型呈现出预测结果“在高发期低估,低发期高估”的现象。6 月份之后的预测结果大大高于实际发病率,其原因在于 6 月份以后,该区为实现全区消除麻疹的目标,出台了一系列的麻疹防控措施,把麻疹防控工作上升到政治的高度,完全遏制了麻疹的发病趋势。这也说明指数平滑法在短期预测中具有一定的效果。

传染病的预测具有很多的难点和不确定性,利用时间序列模型可以使各种因素在时间变量中统一体现并外推。指数平滑法实用范围广,其操作简单,能够大量使用^[5],预测效果与 ARIMA 相近^[13]。麻疹的发病在该区近年来一直处于低发水平,由于麻疹疫苗较为广泛的接种和卫生部门对麻疹疫情的高度关注,大规模暴发的自然因素和社会因素已经不存在,但距离消除麻疹的目标还有较远的距离,也不能排除在局部地区,局部人群中发生小规模暴发,对麻疹疫情的持续监测和预警具有重要的现实意义。

参考文献

[1] 唐雨新,丁愈,任重. 株洲市 2004-2006 年麻疹发病年龄特征分析[J]. 实用预防医学,2007, 14(6):1760-1761.

[2] 杜强,贾丽艳. SPSS 统计分析:从入门到精通[M]. 北京:人民邮电出版社,2009:388-391.

[3] 方积乾. 生物医学研究的统计学方法[M]. 北京:高等教育出版社, 2017:439-442.

[4] 李运明,周厚永,李刚,等. 应用指数平滑法预测医院的人院人次[J]. 西南军医, 2005, 7(1):68-69.

[5] 王昕,程小雯,房师松,等. 指数平滑模型在流感样病例预测中的应用[J]. 中国热带医学,2011, 11(8):938-939.

[6] 郭璐,张敏,朱正平,等. 指数平滑法在南京市艾滋病疫情预测中的应用[J]. 中国艾滋病性病, 2014,20(12):914-916.

[7] 史芸萍,马家奇. 指数平滑法在流行性腮腺炎预测预警中的应用[J]. 中国疫苗和免疫,2010,16(3):233-237.

[8] 张磊,刘艳红. 指数平滑法在预测深圳市宝安区肺结核病人发病人数的应用[J]. 实用预防医学, 2014, 21(8):911-913.

[9] 刘涛,王显军,姜宝法,等. SARIMA 模型预测山东省手足口病发病趋势[J]. 中国卫生统计,2013,30(1):697-700.

[10] 刘罗曼. 时间序列分析中指数平滑法的应用[J]. 沈阳师范大学学报(自然科学版), 2009, 27(4):416-418.

[11] González DA, Sánchez VL, Armas PL, et al. Trends and predictions of tuberculosis, and challenges towards tuberculosis eradication in Havana City[J]. Revista Cubana De Medicina Tropical, 2013, 64(2): 163-175.

[12] 朱奕奕,赵琦,冯玮,等. 应用指数平滑法预测上海市甲型病毒性肝炎发病趋势[J]. 中国卫生统计,2013,30(1):31-35.

[13] 梁静,王培生,李新风,等. 应用 ARIMA 模型对麻疹发病的预测及分析[J]. 实用预防医学, 2016, 23(8):1003-1006.