维普资讯 http://www.cqvip.com

技术与方法 文章编号:0253-3626(2007)03-0275-04

应用灰色预测模型 GM(1,1)对结核病发病率进行预测

易 静,杜昌廷,王润华,刘 琳

(重庆医科大学公共卫生学院卫生统计与信息管理教研室,重庆 400016)

【摘 要】目的:探讨灰色预测模型 GM(1,1)在时间序列资料中的应用;建立结核病登记发病率的预测模型。方法:利用重庆市结核病防治所于 1993~2003 年登记的结核病发病率资料建立灰色预测模型即残差 GM(1,1)模型,并对 2004 年结核病登记发病率进行预测。结果:残差 GM(1,1)模型为:$\hat{X}^{(1)}(k+1)=89.5776e^{0.0392k}-85.0913+\delta(k-1)1.6435e^{-0.3068k}$,1993~2003 年预测值的相对误差均小于 0.05,关联度 > 0.6,小误差概率 $P>0.8$;2004 年结核病登记发病率预测为 5.2942/10 万,相对误差为 0.0461,修正后的相对误差为 0.0121。结论:残差 GM(1,1)模型拟合效果结果较为理想,是一种短期内预测精度较高的预测模型。

【关键词】时间序列;残差 GM(1,1)模型;结核病登记发病率

【中国图书分类法分类号】P49 【文献标识码】A 【收稿日期】2006-09-05

Application of residual grey model(1,1) to the prediction of tuberculosis prevalence

YI Jing,et al

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

【Abstract】*Objective*:To discuss the application of residual Grey Model (1,1) in the time series data to set up the prediction model for Tuberculosis prevalence.*Methods*:Grey Model (1,1)was set up by using the material on Tuberculosis prevalence collected by Tuberculosis Prevention and Treatment Institute of Chongqing city from 1993 to 2003,and then used to predict the Tuberculosis prevalence of 2004.*Results*:The residual Grey Model(1,1)is $\hat{X}^{(1)}(k+1)=89.5776e^{0.0392k}-85.0913+\delta(k-1)1.6435e^{-0.3068k}$, the relative error values of predictive value for 1993 to 2003 were all less than 0.05,and the degree of grey incidence was bigger than 0.6,and the little error probability was bigger than 0.8;The predictive prevalence value in 2004 was 5.2942 per 100 thousands,the relative error was 0.0461 and the modified one was 0.0121.*Conclusion*:The residual grey model (1,1) had a good fitness and can be used to predict tuberculosis prevalence with high prediction precision in short-term time series.

【Key words】Time series;Residual Grey Model(1,1);Tuberculosis prevalence

结核病是严重危害人民群众身体健康和生命安全的重点疾病,近 10 余年来,由于 HIV/AIDS 的流行、移民、多种耐药结核菌的增多,以及不少国家和地区对结核病控制的忽视等因素,全球结核病形势急剧恶化[1,2]。目前,中国属于全球 22 个结核病高负担国家之一,病人数居全球第二,防治形势十分严峻。

灰色模型即 Grey Model,简称 GM(1,1)模型是 1 阶 1 个因素微分方程的时间连续模型[3]。灰色预测通过原始数据的处理和 GM(1,1)模型的建立,发现和掌握系统的发展规律,对系统的未来状态做出科学的定量预测。本研究应用该模型对重庆市结核病防治所登记的 1993~2003 年结核病登记发病率时间序列进行分析,并对 2004 年结核病登记发病率进行预测,为现代结核病控制提供科学依据。

1 资料与方法

1.1 资料来源

资料来源于重庆市结核病防治所 1993~2004 年结核病登记记录及重庆市统计年鉴 1993~2004 年的人口数。

1.2 方法

作者介绍:易 静(1969-),女,副教授,博士生,主要研究方向:人群健康状况评价 。

维普资讯 http://www.cqvip.com

采用 EXCEL 录入人口及结核病登记情况的资料，应用 EXCEL 函数对资料进行分析和预测。

1.3 建立残差 GM(1,1)模型[4]

1.3.1 数据处理

1.3.1.1 列出原始数据序列：$\{x_{(k)}^{(0)}\}(k=1,2,3\cdots\cdots n)$。

1.3.1.2 建立数据向量 Y_n：$Y_n=(x_{(2)}^{(0)},x_{(3)}^{(0)},\cdots\cdots x_{(n)}^{(0)})$。

1.3.1.3 对原始数据序列采用累加的方法构造累加生成列：

$$x_{(k)}^{(1)}=\sum_{i=1}^{K}x_{(i)}^{(0)}(k=1,2,3\cdots\cdots n)。$$

1.3.1.4 利用累加生成列$\{x_{(k)}^{(1)}\}$建立数据矩阵 B：

$$B=\begin{pmatrix}-\frac{1}{2}(x_{(1)}^{(1)}+x_{(2)}^{(1)}) & 1\\ -\frac{1}{2}(x_{(2)}^{(1)}+x_{(3)}^{(1)}) & 1\\ \cdots\cdots\cdots\cdots & \cdots\\ -\frac{1}{2}(x_{(n-1)}^{(1)}+x_{(n)}^{(1)}) & 1\end{pmatrix}$$

1.3.2 建立 GM(1,1)模型

1.3.2.1 模型参数的估计

GM(1,1)模型的微分方程：$\frac{dx^{(1)}}{dt}+ax^{(1)}=\mu$,，其中 a 为发展灰数，μ 为内生控制灰数。设 $a=\begin{pmatrix}a\\ \mu\end{pmatrix}=(B^TB)^{-1}B^TY_n$ 为待估参数向量，利用最小二乘法可求得，即：

$$\hat{X}^{(1)}(k+1)=[X^{(0)}(1)-\frac{\mu}{a}]e^{-ak}+\frac{\mu}{a}(k=1,2,3\cdots\cdots n)$$

1.3.2.2 对 GM(1,1)模型求得的拟合值序列采用累减的方法构造原始数据的预测值序列：

$$\hat{X}_{(k)}^{(0)}=\hat{X}_{(k)}^{(1)}-\hat{X}_{(k-1)}^{(1)}(k=1,2,3\cdots\cdots n)$$

1.3.3 模型检验

1.3.3.1 残差检验

绝对误差：$\Delta_{(i)}^{(0)}=\left|X_{(i)}^{(0)}-\hat{X}_{(i)}^{(0)}\right|(i=1,2,3\cdots\cdots n)$

相对误差：$\varphi_{(i)}=\frac{\Delta_{(i)}^{(0)}}{X_{(i)}^{(0)}}\times 100\%\quad(i=1,2,3\cdots\cdots n)$,

若 $\phi_{(i)}<5\%$，说明拟合和预测理想。

1.3.3.2 关联度检验

关联系数：

$$\eta(k)=\frac{\min\min\left|\hat{X}_{(k)}^{(0)}-X_{(k)}^{(0)}\right|+\rho\max\max\left|\hat{X}_{(k)}^{(0)}-X_{(k)}^{(0)}\right|}{\left|\hat{X}_{(k)}^{(0)}-X_{(k)}^{(0)}\right|+\rho\max\max\left|\hat{X}_{(k)}^{(0)}-X_{(k)}^{(0)}\right|},$$

其中 $\rho=0.5$。

关联度：$r=\frac{1}{n}\sum_{k=1}^{n}\eta(k)>0.60$。

1.3.3.3 后验差检验

原始序列标准差：$S_1=\sqrt{\frac{\sum\left[X_{(i)}^{(0)}-\bar{X}^{(0)}\right]^2}{n=1}}$

绝对误差序列标准差：$S_2=\sqrt{\frac{\sum\left[\Delta_{(i)}^{(0)}-\bar{\Delta}^{(0)}\right]^2}{n-1}}$

小误差概率：$P=p\left\{\left|\Delta_{(i)}^{(0)}-\bar{\Delta}^{(0)}\right|<0.675S_1\right\}>0.80$。

若残差检验、关联度检验和后验差检验都能通过，则可用所建模型进行预测，否则，进行残差修正。

1.3.4 建立残差 GM（1,1）模型 利用 3.1~3.3 步骤建立的 GM(1,1)模型获得的残差序列，按灰色数列模型分折程序(3.1~3.3 步骤)建立相应残差的 GM(1,1)模型：

$\hat{e}^{(1)}(k+1)=\left[e^{(0)}(1)-\frac{\mu_e}{\alpha_e}\right]e^{-\alpha_e k}+\frac{\mu_e}{\alpha_e}$，对$\hat{e}^{(1)}(k+1)$求导数得修正模型为：$\hat{X}^{(1)}(k+1)=\left[X^{(0)}(1)-\frac{\mu}{\alpha}\right]e^{-\alpha k}+\frac{\mu}{\alpha}+\delta(k-1)(-\alpha_e)\left[e^{(0)}(1)-\frac{\mu_e}{\alpha_e}\right]e^{-\alpha_e(k-1)}$，其中 $\delta(k-1)=\begin{Bmatrix}1,k\geqslant 2\\ 0,k<2\end{Bmatrix}$为修正系数。

经残差修正的原始序列预测模型：

$\hat{X}_{(k)}^{(0)}=\hat{X}_{(k)}^{(1)}-\hat{X}_{(k-1)}^{(1)}$,$(k=1,2,3\cdots\cdots n)$。

2 结果与分析

2.1 流行趋势分析

对 1993~2004 年的结核病登记发病率(1/10 万)绘制时序分布图(见图 1)。由图可看出，结核病登记发病率没有明显的起伏变动规律，呈均匀变化现象。

2.2 GM(1,1)预测模型

将原始数据生成累加列$\{x_{(k)}^{(1)}\}$建立数据矩阵 B，

$$B=\begin{pmatrix}-2.24315 & 1\\ -6.49644 & 1\\ \cdots\cdots\cdots & \cdots\\ -48.3923 & 1\end{pmatrix}$$，获得$(B^TB)^{-1}=$

$$\begin{pmatrix}0.000426 & 0.010768\\ 0.010768 & 0.363306\end{pmatrix};$$

由$(B^TB)^{-1}B^TY_n$得到 $\alpha=-0.0392$,$\mu=3.3356$，生

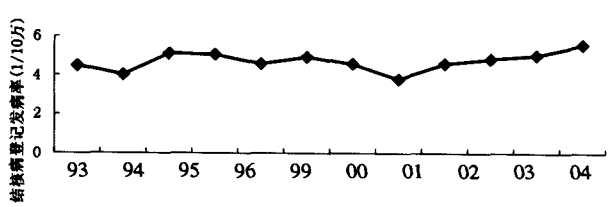


图 1 1993~2004 年结核病登记发病率(1/10 万)时序分布

维普资讯 http://www.cqvip.com

表 1 结核病登记发病率(1/10 万)GM(1,1)分析结果

年度	实测值	预测值	残差值	相对误差	关联系数	修正后的相对误差
1993	4.4863	4.4863	0.0000	0.0000	1.0000	0.0000
1994	4.0203	3.5784	0.4419	0.1099	0.4442	0.0642
1995	5.1082	3.7213	1.3868	0.2714	0.4529	0.0299
1996	5.0502	3.8700	1.1802	0.2337	0.4901	0.0038
1997	4.5642	4.0246	0.5396	0.1182	0.9025	0.0020
1998	4.9219	4.1854	0.7365	0.1496	0.5963	0.0129
1999	4.5631	4.3526	0.2105	0.0461	0.8474	0.0093
2000	3.7842	4.5264	0.7422	0.1961	0.3333	0.0150
2001	4.5642	4.7073	0.1430	0.0314	0.8016	0.0156
2002	4.8259	4.8953	0.0693	0.0144	0.7592	0.0106
2003	5.0078	5.0908	0.0831	0.0166	0.5887	0.0032
2004	5.5505	5.2942	0.2562	0.0461		0.0121
标准差	0.4837	0.2944				

成 GM(1,1)预测模型：

$\hat{X}^{(1)}(k+1)=89.5776e^{0.0392k}-85.0913$。当 k 取不同值时，得到相应的预测值，结果见表 1。

2.3 检验 GM(1,1)预测模型

2.3.1 关联度检验 见表 1，关联度 $r=\frac{1}{n}\sum_{k=1}^{n}\eta(k)=0.6560$，满足 $\rho=5$ 时的检验准则 $r>0.60$。

2.3.2 后验差检验

$P=p\{|\Delta_{(i)}^{(0)}-\overline{\Delta}^{(0)}|<0.675S_1\}=p\{|\Delta_{(i)}^{(0)}-\overline{\Delta}^{(0)}|<0.3263\}=0.8333$，故通过后验差检验。

2.3.3 残差检验 见表 1，比较实测值与相应的预测值对模型进行检验，相对误差最大为 0.2714，可见拟合效果不好，所以要对原模型 $\hat{X}^{(1)}(k+1)=89.5776e^{0.0392k}-85.0913$ 进行残差修正，提高精度。

2.4 残差修正的 GM(1,1)模型

利用残差序列数据，按灰色数列预测模型分折程序，生成残差 GM(1,1)修正模型：$[\hat{e}^{(1)}(k+1)]=1.6435e^{-0.3068k}$。这样经过残差修正后的模型为：$\hat{X}^{(1)}(k+1)=89.5776e^{0.0392k}-85.0913+\delta(k-1)1.6435e^{-0.3068k}$，其中，$\delta(k-1)=\begin{cases}1, k\geq 2\\ 0, k<2\end{cases}$ 为修正系数。修正后的精度提高，修正后的残差计算结果见表 1。

2.5 预测

当 $k=12$ 时，预测出 2004 年结核病的登记发病率为 5.2942，其相对误差为 0.0461，修正后的相对误差为 0.0121。

3 讨 论

结核病发病率预测研究是流行病学研究的重要课题之一，20 世纪 60 年代以来，国内学者对结核病的发病及流行强度进行了大量的预测预报研究。其方法多是六、七十年代的联列递推微分方程组[5]，90 年代的回归模型(直线或曲线模型、指数模型、多项式模型)、指数平滑法[6-9]、ARIMA 模型[10,11]等方法。

早期的联列递推微分方程组常用的 Azuma 模型使用的指标多，未考虑我国结核病控制工作的新进展，如 Azuma 模型设定卡介苗的保护作用为 80%，但现代结核病控制理论认为卡介苗对成人结核病几乎无保护效果等情况，故目前应用很少。

回归模型对于规律性强或接近某一特定曲线类型的序列的预测效果好，能考虑到结核疫情变动的影响因素，保证预测值与观察值的残差最小，不仅可作近期预测，还可作远期预测；但回归模型存在较大缺陷，如要求样本量足够大，样本服从特定的统计分布，在预测时将近期和远期观察值的重要性同等看待，而实际上近期比远期更重要更可靠，因此可能出现量化结果与定性分析结果不符等情况。

指数平滑法对近期观察值比远期观察值给予较大的权数，符合外推的要求，且通过对平滑系数的控制，来提高预测效果的准确性，而它的缺点是需要预先凭经验估计出初始平均值及平滑系数，而这往往要较多的历史资料和进行多次尝试计算。

维普资讯 http://www.cqvip.com

ARIMA 模型是将预测对象随时间推移而形成的数据序列视为一个随机序列，除去个别偶然因素影响观察值外，时间序列是一组依赖于时间 t 的随机变量。这组随机变量所具有的依存关系或自相关性表征了预测对象发展的延续性，而这种自相关性被相应的 ARIMA 模型描述出来，就能从时间序列的过去值、现在值预测其未来的值。ARIMA 模型不需要对时间序列的发展模式作先验假设，方法的本身保证可通过反复识别修改，直至获得满意模型，故 ARIMA 模型适用于各种类型的时间序列数据；在预测时，不仅考察预测变量的过去值与现在值，而且模型拟合产生的误差也作为重要因素进入模型，有利于提高模型的精确度。其缺点是要求较大的样本量，高阶模型中的变量系数在实际工作中无从解释，适合短期预测。

在生物医学的研究中，医学本身提供的信息很多，它可分为三大系统，信息完全明确的称为白色系统，信息完全不明确的称为黑色系统，介于白与黑之间称为灰色系统。灰色理论认为影响结核病发病率的各种因素，一些因素如结核菌素(BCG)的接种水平、综合性医疗机构诊断能力、病人对结核病的认知水平等是可知的，另外一些因素的影响则是未知的，结核病的发展趋势存在白色和黑色参数，由灰色理论可预测结核病发病率的发展趋势。

大多数理统计方法要求大样本并对样本分布有一定要求，要求样本有较好分布规律，且计算工作量大，有时还可能出现定量结果与定性结果不符的现象。灰色系统对样本量及样本分布要求就不那么严格，因此具有适用性强的特点。

重庆市 1993~2004 年结核病登记发病率的原始数据序列无规律可循，且原始数据数量不多。利用残差 GM(1,1)模型对其进行分析与预测，从检验效果看，模型的拟合令人满意，说明残差 GM(1,1)模型具有预测性能好的优点。

(通讯作者：王润华，E-mail：wzhzyp@163.com)

参 考 文 献

[1] Shimao T.Tuberculosis and its control-lessons from the past and future prospect[J].Kekkaku,2005;80(6):481-489.

[2] Aoki M.Tuberculosis control strategy in the 21st century in Japan-for elimination of tuberculosis in Japan[J].Kekkaku,2001;76(7):549-557.

[3] 许汝福，王文昌，易 东，等.时间序列资料 GM(O,N)预测模型及其应用[J].中国卫生统计，1999;16(3):162-163.

[4] 徐国祥，主编.统计预测与决策[M]，第 1 版，上海：上海财经大学出版社，1998.199-225.

[5] 王黎霞，施鸿生，刘胜安，等.我国结核病流行病学模型及疫情态势浅析[J].中国卫生统计，1994;11(4):18-21.

[6] Dye C,Fengzeng Z,Scheele S,et al.Evaluating the impact of tuberculosis control:number of deaths prevented by short-course chemotherapy in China[J].Int J Epidemiol,2000;29(3):558-564.

[7] Nishiura H,Patanarapelert K,Tang IM.Predicting the future trend of drug-resistant tuberculosis in Thailand: assessing the impact of control strategies[J].Southeast Asian J Trop Med Public Health,2004;35(3):649-656.

[8] Rios M,Garcia JM,Sanchez JA,et al.A statistical analysis of the seasonality in pulmonary tuberculosis[J].Eur J Epidemiol,2000;16(5):483-488.

[9] West RW,Thompson JR.Modeling the impact of HIV on the spread of tuberculosis in the United States[J].Math Biosci,1997;143(1):35-60.

[10] Loytonen M,Maasilta P.Multi-drug resistant tuberculosis in Finland-a forecast[J].Soc Sci Med,1998;46(6):695-702.

[11] Antunes JL,Waldman EA.Tuberculosis in the twentieth century:time-series mortality in Sao Paulo,Brazil,1900-97[J].Cad Saude Publica,1999;15(3):463-476.

(上接第 252 页)

[3] 方女画，杨 波，何俏军.乏氧条件下肿瘤细胞的耐药机制及肿瘤治疗策略[J].中国药学杂志，2005;40(16):1208.

[4] Hartmann KA,van der Kleij AJ,Carl UM,et al.Effects of hyperbaric oxygen and normobaric carbogen on the radiation response of the rat rhabdomyosarcoma R1H[J].Int J Radiat Oncol Biol Phys,2001;51(4):1037.

[5] Kunugita N,Kohshi K,Kinoshita Y,et al.Radiotherapy after hyperbaric oxygenation improves radioresponse in experimental tumor models[J].Cancer Lett,2001;164(2):149.

[6] Chendil D,Das A,Dey S,et al.Par-4,a pro-apoptosis gene inhibits radiation-induced NF kappa B activity and Bcl-2 expression leading to induction of radiosensitivity in human prostate cancer cells PC-3[J].Cancer Biol Ther,2002;1(2):152-160.

[7] Kariya S,Ogawa Y,Yoshida S,et al.X-irradiation enhances the expression of Bcl-2 in HL-60 cells:the resulting effects on apoptosis and radiosensitivity[J]. Int J Mol Med,1999;3(2):145-152.

[8] Biroccio A,Candiloro A,Mottolese M,et al.Bcl-2 over-expression and hypoxia synergistically act to modulate vascular endothelial growth factor expression and in vivo angiogenesis in a breast carcinoma line[J].FASEB J,2000;14(5):652.

[9] Sun X,Kanwar JR,Leung E,et al.Gene transfer of antisense hypoxia inducible factor-1 alpha enhance the therapeutic efficacy of cancer immunotherapy[J].Gene Ther,2001;8(8):638.