DOI:10.16016/j.1000-5404.2003.19.036
第25卷第19期 第三军医大学学报 Vol. 25, No. 19
1774 2003年10月 ACTA ACADEMIAE MEDICINAE MILITARIS TERTIAE Oct. 2003

技术方法

文章编号:1000-5404(2003)19-1774-02

多因素时间序列资料 GM(1, N)预测模型及其应用

Predicting model GM(1, N) for multivariate time series data and its application

樊爱军,王开发 (第三军医大学基础医学部数学教研室,重庆 400038)

提 要:目的 对多因素时间序列资料,介绍1阶N个变量的灰色预测模型[Grg Model, GMCL, N],并通过对一实例进行分析来比较3种预测模型[LS估计、GM(0, N)模型和GM(1, N)模型]的优劣。方法 以预测误差平方和和平均模拟相对误差作为评价指标。结果 对同一多因素时间序列资料建立了3种预测模型,并分别计算出它们的预测误差平方和和平均模拟相对误差。结论 对所讨论的多因素时间序列资料,GM系列预测模型的效果优于LS估计,其中GM(1, N)预测模型的效果最优。

对多因素资料预测,一般使用或转化使用最小二乘估计(LS估计)[1]。但对于时间序列资料,LS估计未考虑各因素的发展变化情况,同时由于实验条件或资料收集困难等客观因素,我们所得的时间序列资料只能是小样本,并且观测对象也不一定满足服从某种典型分布,此时再简单地使用LS估计预测可能会带来较大误差。为了克服矛盾,文献[2]介绍了一种时间序列资料灰色GM(0, N)预测模型。本研究将新介绍一种多因素时间序列资料灰色GM(1, N)预测模型,并比较分析3种预测模型对同一多因素时间序列资料实例预测结果的优劣。

1 GM(1, N)预测模型建模方法[3,4]

设 $x_1^{(0)}=\left\{x_1^{(0)}(1), x_1^{(0)}(2), \cdots, x_1^{(0)}(n)\right\}$ 为系统的行为(因子)应变量,$x_i^{(0)}=\left\{x_i^{(0)}(1), x_i^{(0)}(2), \cdots, x_i^{(0)}(n)\right\}$,$i=2, 3, \cdots, N$ 为系统的行为(因子)自变量,$x_i^{(0)}$,$i=1, 2, \cdots, N$ 均为时间序列资料。

记 $x_i^{(1)}(k)=\sum_{j=1}^{k} x_i^{(0)}(j)$, $k=1, 2, \cdots, n$

为 $x_i^{(0)}$ 的1-AGO序列,$i=1, 2, \cdots, N$。再记

$$x_1^{(o)}(k)=0.5x_1^{(1)}(K)+0.5x_1^{(1)}(K-1), \quad k=2, 3, \cdots, n$$

为 $x_1^{(1)}$ 的紧邻均值生成序列。那么GM(1, N)模型的灰微分方程形式为

$$x_1^{(o)}(k)+\alpha x_1^{(1)}(K)=\sum_{i=2}^{N} b_i x_i^{(1)}(k)$$

其白化方程为

$$\frac{dx_1^{(1)}}{dt}+\alpha x_1^{(1)}=\sum_{i=2}^{N} b_i x_i^{(1)}$$

其中参数列â为:

$$\hat{a}=[a, b_2, \cdots, b_N]^T=(B^TB)^{-1}B^TY_N$$

$$B=\begin{bmatrix} -z_1^{(1)}(2) & x_2^{(1)}(2) & \cdots & x_N^{(1)}(2) \\ -z_1^{(1)}(3) & x_2^{(1)}(3) & \cdots & x_N^{(1)}(3) \\ \cdots & \cdots & \cdots & \\ -z_1^{(1)}(n) & x_2^{(1)}(n) & \cdots & x_N^{(1)}(n) \end{bmatrix}$$

$$Y_N=[x_1^{(0)}(2), x_1^{(0)}(3), \cdots, x_1^{(0)}(n)]^T$$

一般地,按

$$x_1^{(0)}(k)=\sum_{i=2}^{N}\beta_i x_i^{(1)}(k)-\alpha x_1^{(1)}(k-1)$$

检验模型,其中 $\alpha=\frac{a}{1+0.5a}$,$\beta_i=\frac{b_i}{1+0.5a}$,$i=2, 3, \cdots, N$。

2 实例分析

数据资料来源于文献[5]。收集了1984—1997年流行性出血热监测区流行性出血热(EHF)的年发病率($x_1^{(0)}$),当年一季度平均鼠密度($x_2^{(0)}$),去年第四季度平均人群阴性感染率($x_3^{(0)}$)。下面我们以1984—1993年的数据建立模型,对1994—1997年该地区EHF的年发病率进行预测。

2.1 三种模型的建立

首先计算序列 $x_i^{(0)}$ 的1-AGO序列 $x_i^{(1)}$。

按照上述方法建立1985—1993年的GM(1, N)预测模型为:

$$\hat{x}_1^{(0)}(k)=5.0553x_2^{(1)}(K)-0.8067x_3^{(1)}(k)-0.3856x_1^{(1)}(k-1)$$

按照文献[2]的方法建立1985—1993年的GM(0, N)预测模型为:

$$\hat{x}_1^{(0)}(k)=11.5626+24.5214x_2^{(1)}(k)-7.92584x_3^{(1)}(k)-x_1^{(1)}(k-1)$$

应用LS估计建立1984—1993年的回归方程为:

$$\hat{x}_1=-1.6979782+18.295931x_2-4.71959x_3$$

2.2 三种模型的预测比较分析

对1985—1997年该地区EHF年发病率的预测结果见表1。

作者简介:樊爱军(1971—),女,湖南省安仁县人,讲师,主要从事生物数学方面的研究。电话:(023)68752211

收稿日期:2002-05-09;修回日期:2002-12-20

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

GM(1, N)模型的预测误差平方和($\sum_{i=1}^{n}\varepsilon_i^2$)为 914.996 7, 平均模拟相对误差($\frac{1}{n}\sum_{i=1}^{n}\Delta_i$)为 0.284 9, GM(0, N)模型的预测误差平方和为 1 690.369 9, 平均模拟相对误差为 0.291 8, LS 估计的预测误差平方和为 2 970.008 3, 平均模拟相对误差为 0.297 3。因此, 对 1985—1997 年该地区 EHF 年发病率而言, GM 系列模型的预测效果优于 LS 估计的预测效果, 其中 GM(1, N)模型的预测效果最为满意。

表 1　三种模型对某地区 EHF 年发病率的预测结果

年份	实际值	预测值			预测误差(ε_i)			相对误差(Δ_i)		
		GM(1, N)	GM(0, N)	LS	GM(1, N)	GM(0, N)	LS	GM(1, N)	GM(0, N)	LS
1985	13.02	11.307 0	13.292 2	14.690 3	−1.713 0	0.272 2	1.670 3	0.131 6	0.020 9	0.128 3
1986	16.06	14.390 5	14.261 1	16.132 6	−1.669 5	−1.798 9	0.072 6	0.103 9	0.112 0	0.004 5
1987	14.05	16.333 6	13.507 5	16.776 0	2.283 6	−0.542 5	2.726 0	0.162 5	0.038 6	0.194 0
1988	22.98	21.554 1	25.016 0	24.975 9	−1.425 9	2.036 0	1.995 9	0.062 0	0.088 6	0.086 9
1989	20.90	21.983 4	22.903 1	20.936 6	1.083 4	2.003 1	0.036 6	0.051 8	0.095 8	0.001 8
1990	22.01	22.873 8	19.772 0	19.046 5	0.863 8	−2.238 0	−2.963 5	0.039 2	0.101 7	0.134 6
1991	24.00	25.362 7	24.878 0	26.161 8	1.362 7	0.878 0	2.161 8	0.056 8	0.036 6	0.090 1
1992	28.01	27.519 8	28.503 0	27.009 4	−0.490 2	0.493 0	−1.000 6	0.017 5	0.017 6	0.035 7
1993	27.99	26.803 4	26.895 0	24.574 9	−1.186 6	−1.095 0	3.415 1	0.042 4	0.039 1	0.122 0
1994	27.95	27.095 0	35.136 0	30.454 9	−0.855 0	7.176 0	2.494 9	0.031 1	0.256 7	0.089 2
1995	11.15	23.867 5	11.233 0	10.843 0	12.717 5	0.083 0	−0.307 0	1.140 6	0.007 4	0.027 5
1996	13.52	34.516 1	48.875 0	41.631 2	20.996 1	35.355 0	28.111 2	1.553 0	2.615 0	2.079 2
1997	52.97	35.844 0	33.737 0	6.784 3	−17.126 0	−19.233 0	−46.185 7	0.323 3	0.363 1	0.871 5

3　讨论

概率统计和灰色系统理论都是常用的处理不确定系统的研究方法, 其中概率统计研究的是随机不确定现象, 着重于考察随机不确定现象的历史统计规律, 其出发点是大样本, 并要求对象服从某种典型分布, 而灰色系统理论则着重研究解决小样本、贫信息不确定问题, 它依据信息覆盖, 通过序列生成来寻求现实规律, 其特点是少数据建模[4]。因此, 对于小样本的时间序列资料, 由于 LS 估计直接利用原始数据来进行计算与预测, 未考虑各因素的发展变化的情况, 而 GM 系列模型则是利用原始数据序列 $x_i^{(0)}$ 的 1-AGO 序列 $x_i^{(1)}$ 来进行计算与预测, 由于 AGO 序列能弱化随机干扰, 比原始序列更能体现事物发展规律, 并且 GM 系列模型对样本含量和概率分布都无严格要求, 所以此时选用 GM 系列模型能更好地考虑各时间序列资料的发展变化情况, 其预测效果更加满意。再进一步分析 GM 系列模型知道, 由于 GM(0, N)模型不含导数, 为静态模型, 而 GM(1, N)模型含有灰导数, 是动态模型, 因此 GM(1, N)模型的预测效果最为满意。但对于大样本数据资料, 由于此时对象的分布具有典型性, 所以选用 LS 估计更为合理。

需要说明的是本研究所用 3 种预测模型都只是根据当前若干个序列建立的预测模型, 因此它反映的也只是事物近期的发展变化规律, 可用它们来进行短期预测。但若用于中长期预测则是不太适合的(如表 1 的结果所示)。如何确定各预测模型的有效超前期呢? 3 种模型有效超前期的优劣又怎样呢? 文献[5] 关于 LS 估计的有效超前期作了讨论, 关于 GM 系列模型的讨论我们将另文进一步分析。

关键词: 多因素时间序列; 预测模型; 预测误差平方和; 平均模拟相对误差

中图法分类号: R311　　文献标识码: B

参考文献:

[1] 郭祖超. 医用数理统计方法[M]. 北京: 人民卫生出版社, 1987. 377—381.

[2] 许汝福, 王文昌, 易　东, 等. 时间序列资料 GM(0, N)预测模型及其应用[J]. 中国卫生统计, 1999, 16(3): 162—163.

[3] 邓聚龙. 灰色控制系统[M]. 武汉: 华中理工大学出版社, 1997. 346—365.

[4] 刘思峰, 郭天榜, 党耀国, 等. 灰色系统理论及其应用[M]. 北京: 科学出版社, 1999. 1—4, 123—126.

[5] 薛付忠, 王洁贞, 马希兰. 线性回归预测模型有效超前期的确定方法[J]. 中国卫生统计, 1999, 16(3): 155—156.

(编辑　张　宁)

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net