

基于 EXCEL 建立人口灰色预测模型的研究

——以天津市农业人口为例

常淑玲^{1,2} (1. 天津大学环境科学与工程学院, 天津 300072; 2. 西藏自治区拉萨市环境保护局, 西藏拉萨 850000)

摘要 以 EXCEL 为基础, 建立人口 GM(1,1) 预测模型, 通过在编制好的 Excd 表格中输入已知时刻的实际值, 就可以预测出所需的结果。以天津市农业人口为例, 并检验精度, 取得了较好的预测结果, 为人口预测提供了一种操作简单、精度较高的方法。

关键词 GM(1,1); EXCEL; 农业人口; 预测

中图分类号 F323.6 文献标识码 A 文章编号 0517 - 6611(2007) 09 - 02825 - 02

Research on the Establishment of Grey Sequence Forecast Model based on Excd

CHANG Shu-ling (Schol of Environnert Science and Technology, Tianjin University, Tianjin 300072)

Abstract The prediction of population is necessary in the plan of city, the social development index and prediction of water demand. In this paper, we established GM(1,1) model with Excel. We forecasted agricultural population of Tianjin and estimated precision. The consequence was satisfied.

Key words GM(1,1); EXCEL; Agricultural Population; Forecast

1 人口预测模型

人口预测方法很多, 主要有一元线性回归法、人口自然增长法、马尔萨斯指数增长法、指数平滑法、指数增长法、Logistic 法、宋健模型法和 GM(1,1) 法等。不同的方法具有不同的适用范围和特点: 一元线性回归法适用于数据直线趋势较明显的预测; 自然增长率等资料准确可靠时, 可采用人口自然增长法或马尔萨斯法; 历史数据较少时, 可采用指数平滑和移动平均数法; 宋健模型法用于短期预测的精度较高; 数据情况复杂并暗藏指数规律时可采用 GM(1,1) 模型法^[1]。分析天津市 1995 ~ 2005 年人口资料, 呈现明显的指数规律, 可采用 GM(1,1) 模型。

2 GM(1,1) 预测模型

2.1 模型的建立 GM(1,1) 模型为单序列的一阶线性动态模型, 是常用的一种灰色数列预测模型, 构造如下所述。

对于给定的原始时间数据序列: $x^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)\}$; 经过累加生成新的数据序列: $x^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)\}$, 其中 $x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i)$, $k = 1, 2, \dots, n$ 。作均值序列: $Z^{(1)}(k) = 0.5x^{(1)}(k-1) + 0.5x^{(1)}(k)$ 。构造数据矩阵 B:

$$B = \begin{bmatrix} Z^{(1)}(2) & 1 & -0.5[x^{(1)}(1) + x^{(1)}(2)] & 1 \\ Z^{(1)}(3) & 1 & -0.5[x^{(1)}(2) + x^{(1)}(3)] & 1 \\ \dots & \dots & \dots & \dots \\ Z^{(1)}(n) & 1 & -0.5[x^{(1)}(n-1) + x^{(1)}(n)] & 1 \end{bmatrix}$$

构造数据向量: $Y_N = [x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n)]^T$; 则

白化形式的微分方程为: $\frac{dx^{(1)}}{dt} + ax^{(1)} = u$, 式中 a, u 为待估计的参数; 辨识微分方程参数: $\hat{a} = (B^T B)^{-1} B^T Y_N = [a, u]^T$;

得时间序列预测模型: $X^{(1)}(k+1) = x^{(0)}(1) - u/a e^{-ak} + u/a$, 即为一阶累加数列的拟合模型; 预测数据还原即为预测值: $X^{(0)}(k) = X^{(1)}(k) - X^{(1)}(k-1)$ 。

2.2 模型精度的检验 灰色系统理论一般采用 3 种方式检验来判断模型的精度。

(1) 残差大小检验: 对模型值和实际值的误差进行逐点检验, 是一种直观的逐点进行比较的算术检验。

定义残差 $\epsilon(k) = x^{(0)}(k) - X^{(0)}(k)$, 则残差序列 $\epsilon^{(0)} = [\epsilon(1), \epsilon(2), \dots, \epsilon(n)]$, 定义 $r_k = \left| \frac{\epsilon(k)}{x^{(0)}(k)} \right|$, 平均模型相对误差为 $\bar{r} = \frac{1}{n} \sum_{k=1}^n r_k$, 给定 α , 当 $\bar{r} < \alpha$ 成立时, 称模型为残差合格模型。

(2) 后验差检验: 对残差分布的统计特征进行检验。 $x^{(0)}$ 的均值与方差为: $\bar{x} = \frac{1}{n} \sum_{k=1}^n x^{(0)}(k)$, $S_1^2 = \frac{1}{n} \sum_{k=1}^n [x^{(0)}(k) - \bar{x}]^2$ 。 $\epsilon^{(0)}$ 的均值与方差为: $\bar{\epsilon} = \frac{1}{n} \sum_{k=1}^n \epsilon^{(0)}(k)$, $S_2^2 = \frac{1}{n} \sum_{k=1}^n [\epsilon^{(0)}(k) - \bar{\epsilon}]^2$ 。

称 $C = S_2/S_1$ 为均方差比值。对于给定的 C_0 , 当 $C < C_0$ 模型时, 称模型为均方差合格模型; $p = P[|\epsilon(k) - \bar{\epsilon}| < 0.6745 S_1]$ 称为小误差概率。对于给定的 p_0 , 当 $p > p_0$, 称模型为小误差概率合格模型。

(3) 关联度检验: 通过考察模型曲线与建模序列曲线的相似程度进行检验, 属几何检验。

$x^{(0)}$ 的关联度为:

$$|s| = \left| \frac{1}{n-1} \sum_{k=2}^n [x^{(0)}(k) - x^{(0)}(1)] + \frac{1}{2} [x^{(0)}(n) - x^{(0)}(1)] \right|$$

$X^{(0)}$ 的关联度为:

$$|S| = \left| \frac{1}{n-1} \sum_{k=2}^n [X^{(0)}(k) - X^{(0)}(1)] + \frac{1}{2} [X^{(0)}(n) - X^{(0)}(1)] \right|$$

关联度 $r = \frac{1 + |s| + |S|}{1 + |s| + |S| + |S - s|}$, 若对于给定的 r_0 , 有 $r > r_0$, 则称模型为关联度合格模型^[3]。

3 天津市农业人口预测

3.1 概况 天津市为中国 4 大直辖市之一, 是中国北方最大的沿海开放城市, 环渤海地区的经济中心, 地处华北平原东部, 属暖温带半湿润大陆季风型气候, 总面积 1.19 万 km²。现辖 18 个辖区县。天津市多年平均降水量为 492 mm, 人均水资源量 101.92 m³(2005 年), 每年依靠引滦入津和引黄济津跨流域调水措施以缓解天津市的用水危机。生活用水是天津市用水单元的主要部分之一, 要预测人口用水量, 则用水的主体——人口预测的准确度就尤为重要。

2005 年, 天津市户籍人口 939.31 万, 其中农业人口

作者简介 常淑玲(1978-), 女, 河北唐山人, 在读硕士, 助理工程师, 从事水资源可持续发展模型研究。

收稿日期 2006-12-25

376.95万,人口自然增长率1.43‰。以天津市1995~2005年的农业人口数据为基础数据,用EXCEL建立灰色预测模型并检验其精度。

3.2 人口GM(1,1)模型与检验 以天津市农业人口为例,基于Excel的矩阵计算功能,建立GM(1,1)预测模型并进行精度检验。分析天津市1995~2005年的统计数据,呈现明显的指数趋势,笔者选取GM(1,1)法,并运用常用的办公软件EXCEL建立人口GM(1,1)。

数据见表1,建立农业人口GM(1,1)模型及精度检验,数据输入及操作如下:将1995~2005年农业人口数据填入C7:C17;复制C7到D7和E7,D8输入“D8=D7+C8”,用填充柄填充D9:D17;C24中输入“=0.5*(D7+D8)”,D24输入1,用填充柄向下填充C25:D33;选定区域C34:L35,输入“=TRANSPOSE(C24:D33)”;选定区域C36:D37,输入“=MMULT(C34:

L35,C24:D33)”;选定区域H6:I37,输入“=MINVERSE(C36:D37)”;选定区域C38:L39,输入“=MMULT(H6:I37,C34:L35)”;选定区域C40:L41,输入“=MMULT(C38:L39,C8:L17)”;H40输入“=C41/C40”;E8中输入“=-0.5*(C40*(C7-H40)*(EXP(-C40*B7)+EXP(-C40*B8))+2*(H40)+C41)”,用填充柄向下填充‘E9:E17’(矩阵运算,如TRANSPOSE,MMULT,MINVERSE,输入相应公式后,按“Ctrl+Shift+Enter”组合键)。精度检验部分的计算都是EXCEL的基本运算^[4],这里不再赘述。

农业人口精度检验:平均相对误差 $\rho = 0.0026 < 0.01$,精度为一级(好);关联度 $\gamma = 0.9270 > 0.90$,精度为一级(好);均方差比值 $C = 0.0606 < 0.65$,精度为三级(勉强);小误差概率 $p = 0.9090 > 0.90$,精度一级(好)。

可见,用EXCEL建立的GM(1,1)农业人口精度很高,模

表1

农业人口GM(1,1)及精度检验

(B) 序号	(A) 年份	(C) 农业人口 万人	(D) 累加序列	(E) 预测数列	残差检验		后验差检验			关联度检验	
					(F) 残差序列	(G) 相对误差序列	(H) $x^{(0)}$ 方差平方	(I) $(^{(0)})$ 方差平方	(J) 小误差概率	(K) s	(L) S
1	1995	386.77	386.77	386.77	0.0000	0.0000	37.1016	0.9893	0.0000		
2	1996	385.26	772.02	383.91	1.3487	0.0035	20.9821	0.1253	1.3559	-1.5105	-2.8592
3	1997	384.48	1156.51	382.80	1.6830	0.0044	14.5146	0.4738	1.6921	-2.2813	-3.9643
4	1998	383.76	1540.26	381.70	2.0586	0.0054	9.5075	1.1320	2.0697	-3.0077	-5.0663
5	1999	381.45	1921.72	380.60	0.8515	0.0022	0.6045	0.0205	0.8561	-5.3136	-6.1651
6	2000	379.48	2301.20	379.51	-0.0219	0.0001	1.4198	1.0335	0.0221	-7.2826	-7.2607
7	2001	378.75	2679.95	378.41	0.3406	0.0009	3.6919	0.4277	0.3425	-8.0125	-8.3532
8	2002	377.91	3057.87	377.32	0.5900	0.0016	7.6252	0.1637	0.5932	-8.8525	-9.4425
9	2003	376.23	3434.10	376.24	-0.0034	0.0000	19.7219	0.9960	0.0034	-10.5320	-10.5287
10	2004	376.38	3810.48	375.15	1.2231	0.0032	18.4690	0.0522	1.2297	-10.3887	-11.6118
11	2005	376.95	4187.42	374.07	2.8710	0.0076	13.9102	3.5205	2.8864	-9.8207	-12.6917

型是合格的。笔者预测至2015年,天津市农业人口为363.44万人,比1995年减少了6%,说明天津市的城镇化率很高。

4 讨论

(1) 用EXCEL建立的GM(1,1)人口预测模型是一张动态的电子表格,只要具有相同的样本个数,就可直接应用,即使不懂编程的人员也可以轻松掌握和使用GM(1,1),大大降低了对工作人员的技术要求。

(2) GM(1,1)适用于具有潜在指数发展规律的数据,不仅适用于预测农业人口,还可以预测灾害发生的概率、城市需

水量、大气质量、土壤中农药残留等,实用性很强。

参考文献

- [1] 汤江龙,赵小敏.土地利用规划中人口预测模型比较研究[J].中国土地科学,2005,19(2):14-20.
- [2] 邓聚龙.灰色系统基本方法[M].武汉:华中理工大学出版社,1987:12-14.
- [3] 郭齐胜,杨秀月.系统建模与仿真及其军事应用系列丛书——系统建模[M].北京:国防工业出版社,2006:223-240.
- [4] 唐五湘,程桂枝.Excel在预测中的应用[M].北京:电子工业出版社,2001.