安徽科技学院学报,2008,22(6):73~76
Journal of Anhui Science and Technology University

中国人口增长预测模型

冯守平

(安徽财经大学　统计与应用数学学院,安徽　蚌埠　233030)

摘　要:本文以 Logistic 人口阻滞增长模型为基础建立了我国人口增长的预测模型,对各年份全国人口总量增长的短、中、长期趋势作出了预测。从而为我国人口的控制与管理提供了一定的依据。

关键词: Logistic 模型;最小二乘法;人口增长;Excel 软件;MATLAB 软件

中图分类号:O213　　**文献标识码**:A　　**文章编号**:1673-8772(2008)06-0073-04

On Prediction Model of China Population Growth

FENG Shou-ping

(School of Statistics & Applied Mathematics, Anhui University of Finance and Economics, Bengbu 233030, China)

Abstract: By retarding increase model, this article gives prediction on China population growth model and on short-term, medium-term and long-term trend of China population growth; it provides a basis for the control and management of China population.

Key words: Population growth; Logistic model; Least squares methed; Excel; LINGO10

中国是一个人口大国,人口问题始终是制约我国经济发展的关键因素之一。我国人口发展经历了多个阶段,近年来的人口发展出现了一些新的特点,例如,老龄化进程加速,出生人口性别比持续升高,以及乡村人口城镇化等因素,这些都影响着中国人口的增长。全面建设小康社会时期是我国社会快速转型期,人口发展面临着前所未有的复杂局面,人口安全面临的风险依然存在。因此,如何准确地判断我国人口在未来若干年的发展趋势就显得非常重要。

1　Logistic 阻滞增长模型的原理

Logistic 阻滞增长模型是荷兰生物数学家 Verhulst19 世纪中叶提出的,它不仅能够较好地描述人口与许多生物数量的变化规律,而且在经济领域也有广泛的应用。阻滞增长模型是考虑到自然资源、环境条件等因素对人口增长的阻滞作用,并且随着人口的增加,阻滞作用越来越大,所谓阻滞增长模型就是对指数增长模型的基本假设进行修改后得到的。阻滞作用体现在对人口增长率的影响上,使得 r 随着人口数量 x 的增加而下降。若将 r 表示为 x 的函数 r(x)。则它应是减函数[1]。于是有:

$$\frac{dx}{dt}=r(x)x,x(0)=x_0, \tag{1}$$

对 r(x)的一个最简单的假定是,设 r(x)为 x 的线性函数,即

$$r(x)=r-sx(r>0,s>0), \tag{2}$$

其中,r 称为固有增长率,表示人口很少时,即 $x\approx0$ 时的增长率。设自然资源和环境条件所能容纳的

收稿日期:2008-09-15

基金项目:安徽省教育厅处自然科学基金项目(2006KJ052C)。

作者简介:冯守平(1954-),男,安徽省淮南市人,学士,副教授,主要从事数理统计研究。

最大人口数量 x_m，称为人口容量。当 $x=x_m$ 时人口不再增长，即增长率 $r(x_m)=0$，代入(2)式得 $s=\frac{r}{x_m}$，于是(2)式变为 $r(x)=r(1-\frac{x}{x_m})$，将它代入方程(1)得：

$$\begin{cases}\frac{dx}{dt}=rx(1-\frac{x}{x_m}),\\ x(0)=x_0\end{cases} \tag{3}$$

解方程(3)可得：

$$x(t)=\frac{x_m}{1+(\frac{x_m}{x_0}-1)e^{-rt}} \tag{4}$$

这就是我们下面要用的 Logistic 方程。

2 人口模型的建立、分析与预测

为了对我国今后的人口总数做出预测，我们首先从中国经济统计数据库(http://211.86.245.155/index.aspx)上查到我国从 1954 年到 2005 年全国总人口的数据如表 1 所示：

表 1 各年份全国总人口数(单位：千万)

年份	总人口	年份	总人口	年份	总人口	年份	总人口	年份	总人口
1954	60.2	1965	72.5	1976	93.7	1987	109.3	1998	124.761
1955	61.5	1966	74.5	1977	95.0	1988	111.026	1999	125.786
1956	62.8	1967	76.3	1978	96.259	1989	112.704	2000	126.743
1957	64.6	1968	78.5	1979	97.5	1990	114.333	2001	127.627
1958	66.0	1969	80.7	1980	98.705	1991	115.823	2002	128.453
1959	67.2	1970	83.0	1981	100.1	1992	117.171	2003	129.227
1960	66.2	1971	85.2	1982	101.654	1993	118.517	2004	129.988
1961	65.9	1972	87.1	1983	103.008	1994	119.850	2005	130.756
1962	67.3	1973	89.2	1984	104.357	1995	121.121		
1963	69.1	1974	90.9	1985	105.851	1996	122.389		
1964	70.4	1975	92.4	1986	107.5	1997	123.626		

2.1 将 1954 年看成初始时刻，即 $t=0$，1955 为 $t=1$，以此类推，以 2005 年为 $t=51$ 作为终时刻。用函数(4)对表 1 中的数据进行非线性拟合，运用 Matlab 编程[2]，得到相应的参数 $x_m=180.9871$，$r=0.0336$，又 $x_0=62.2$，所以，相应的回归方程为

$$x(t)=\frac{180.9871}{1+(\frac{180.9871}{60.2}-1)e^{-0.0336t}} \tag{5}$$

同时算出可决系数(可决系数是判别曲线拟合效果的一个重要指标)：

$$R^2=1-\sum_{i=1}^{5}(y_i-\hat{y}_i)^2/\sum_{i=1}^{5}(y_i-\bar{y})^2=0.9959$$

结果分析：从表 1 所给信息可知从 1951～1958 年为我国第一次出生人口高峰，形成了中国人口的快速增长期；1959、1960、1961 年为三年自然灾害时期，这段时期人口的增长受到很大影响，变成了负增长，因此这段时期人口波动较大，总的来说，1951～1962 年的人口增长的过程并不仅仅受到很多小的随机因素影响，这从下面的图 1 中最长的一条曲线形状可以看出。因此模型的随机误差不是服从正态分布。另外由可决系数来看，拟合的效果虽然比较理想，但这只是表示用函数(5)拟合 1954～2005 年的总人口数据的效果比较理想，并不表示用这个回归方程来作我国总人口预测时的效果也好，因为 2005 年以后的人口发展的环境毕竟与 50 年代、60 年代、70 年代有着很大的区别。

由于上面使用函数(4)去拟合表 1 中的数据时，其拟合的过程实际上需要使用最小二乘法，而模型的随机误差并不服从正态分布，因此这种拟合的合理性值得怀疑。因此我们再选择 1963 年作为初始年，对表 1 中的数据进行拟合。

2.2 将 1963 年看成初始时刻，即 $t=0$，以 2005 年为 $t=32$ 作为终时刻。用函数(4)对表 1 中的数据进行非线性拟合，运用 Matlab 编程，得到相应的参数 $x_m=151.4513$，$r=0.0484$ 又 $x_0=69.1$，所以，相应的回归

方程为：
$$x(t)=\frac{151.4513}{1+(\frac{151.4513}{69.1}-1)e^{-0.0484t}} \tag{6}$$

相应的可决系数 $R^2=0.9994$。

结果分析：1963～1979年期间，人口的增长在城市与农村按照各自的规律增长，农村基本上按照自然方式增长，而城市特别是人口相对密集的大、中城市，由于受到经济收入、文化、医疗等方面的影响，生育率得到相当的控制。城市与农村人口增长的方式有很大的不同，这种差别并不是一些小的随机因素，这从下面的图1中较长的一条曲线形状可以看出。总的来说，可以认为这一阶段随机误差仍然不服从正态分布。因此用函数(4)去拟合表1中的数据时，其合理性受到质疑。另一方面，2005年以后的人口发展的环境毕竟与60年代、70年代有着很大的区别，而此次回归包含了太多的60年代、70年代信息，所得到的回归方程能否用来预测今后若干年我国人口发展趋势，还有待作进一步的分析。因此我们再选择1980年作为初始时刻，对表1中的数据进行拟合。

2.3　将1980年作为初始刻，即 $t=0$，以2005年，即 $t=25$ 作为终时刻，用函数(4)对表1中的数据进行非线性拟合，运用Matlab编程，得到相关的参数 $x_m=153.5351$，$r=0.0477$，又 $x_0=98.705$ 相应的回归方程为：
$$x(t)=\frac{153.5351}{1+(\frac{153.5351}{98.705}-1)e^{-0.0477t}} \tag{7}$$

相应的可决系数为：$R^2=0.9987$。

结果分析：从1980～2005年，人口增长方式与上述两个阶段都不同：国家计划生育政策逐渐得到完善及全面贯彻落实，这个时期的人口增长，无论是农村还是城市，都受到国家计划生育政策的严格控制，而且计划生育政策在这一时期是相当稳定的，这一时期影响人口增长的其他因素基本上都可以看做是小的随机的，因此我们可以认为这一阶段人口增长模型的随机误差近似服从正态分布，这从下面图1中较短的曲线形状可以看出。另外，根据《国家人口发展战略研究报告》，今后我国的总和生育率将继续控制在1.8的水平上，这与80、90年代水平一致，2005年以后的人口发展环境整体上与80、90年代极为相似，因此我们选择1980年作为初始年份，2005年作为终时刻进行拟合，用所得回归方程(7)为今后人口发展进行预测，应该更加可信。

利用Excel软件，作出利用回归方程(5)、(6)、(7)作预测时的残差形成的残差图[3]，如图1所示：

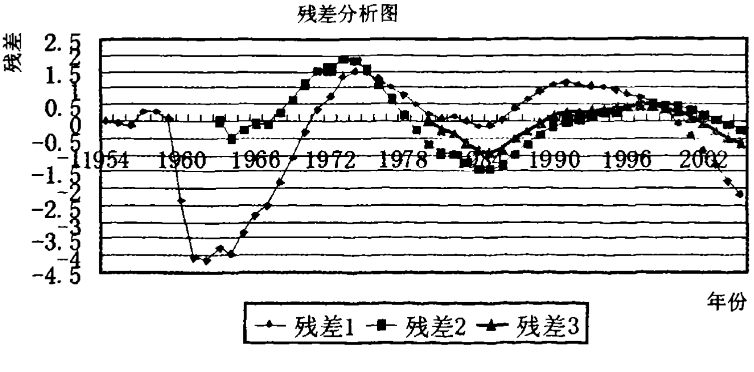

图1　用三个回归方程预测时形成的残差折线图

我们分别根据三个不同的回归方程(5)、(6)、(7)，对各年份总人口进行预测，得到结果见表2：

表 2 各年份全国总人口用不同回归方程预测总人口数(单位:千万)

年份	预测值(5)	预测值(6)	预测值(7)	年份	预测值(5)	预测值(6)	预测值(7)
2000	126.7649	126.3338	126.473	2027	154.3392	143.7168	144.9778
2003	130.5141	129.2303	129.5168	2030	156.5494	144.7157	146.0632
2006	134.1	131.8447	132.2758	2033	158.6028	145.5908	147.0172
2009	137.516	134.1926	134.7638	2036	160.5063	146.3562	147.8541
2012	140.7577	136.2917	136.9971	2039	162.267	147.0247	148.5871
2015	143.8231	138.1607	138.9933	2042	163.8924	147.6077	149.2284
2018	146.7117	139.819	140.771	2045	165.3903	148.1158	149.7886
2021	149.4251	141.2856	142.3489	2048	166.7683	148.558	150.2775
2024	151.9662	142.579	143.7452				

根据三个不同的回归方程(5)、(6)、(7),利用 Excel 软件作出它们相应的图形,得下面图 2:

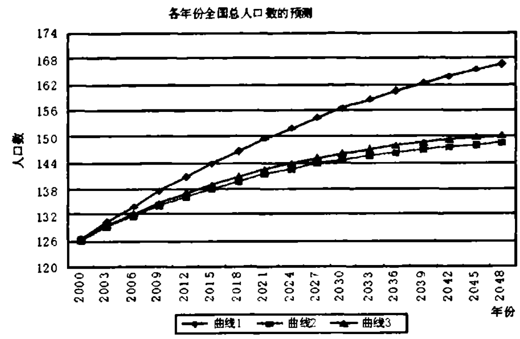

图 2 各年份全国总人口数预测值的折线图

根据国家人口发展战略研究课题组 2007 年所发表的《国家人口发展战略研究报告》,我国的人口 2010 年控制在 13.6 亿,2020 年控制在 14.5 亿,2050 年前后达到高峰 15 亿左右.由上表可以看出:用回归方程(5)预测得到的数据偏大,在 2024 年总人口就已经超过了 15.19662,而且一直以比较快的速度增长到 2048 年达 1.667683 亿,且当 $t\to\infty$ 时,$x(t)\to x_m=18.09871$ 亿,其人口峰值过大;用回归方程(6)预测得到的数据偏小,2010 年,2020 年,2050 年的三个人口预测值都偏小;用回归方程(7)预测的数据总体来讲更接近《国家人口发展战略研究报告》的预测值,且当 $t\to\infty$ 时,$x(t)\to x_m=15.35351$ 亿,此为人口发展的峰值。上面图 2 中的曲线 3 正是我们所要的。

3 结束语

本文仅仅讨论了我国总人口数在未来一段时间内的发展趋势,未涉及其他的指标,实为挂一漏万,因为人口发展的综合研究是一个非常复杂的系统工程,非一篇两篇论文所能完成。

另外,根据本文的讨论,我们可以得出如下的结论:(1)在对一组数据进行分析、回归、预测时,在一定的情形下,并不是数据用的越多越好,重要的是所用的数据与我们要预测的数据产生的背景是否相似;(2)若涉及到最小二乘法时,模型的随机误差是否服从正态分布,因为只有当模型的随机误差从正态分布时,最小二乘法才具有一些优良的性质。

参考文献:

[1]姜启源,谢金星,叶俊.数学模型[M].北京:高等教育出版社,2003.

[2]张志勇.精通 MATLAB6.5 版[M].北京:航空航天大学出版社,2004.

[3]于洪彦.Excel 统计分析与决策[M].北京:高等教育出版社,2006.

(责任编辑:宾 鹏)