## 2007 高教社杯全国大学生数学建模竞赛

## 承 诺 书

我们仔细阅读了中国大学生数学建模竞赛的竞赛规则.

我们完全明白,在竞赛开始后参赛队员不能以任何方式(包括电话、电子邮件、网上咨询等)与队外的任何人(包括指导教师)研究、讨论与赛题有关的问题。

我们知道,抄袭别人的成果是违反竞赛规则的,如果引用别人的成果或其他 公开的资料(包括网上查到的资料),必须按照规定的参考文献的表述方式在正 文引用处和参考文献中明确列出。

我们郑重承诺,严格遵守竞赛规则,以保证竞赛的公正、公平性。如有违反 竞赛规则的行为,我们将受到严肃处理。

我们参赛选择的题号是(从 A/B/C/D 中选择一项填写):A							
我们的参赛报名号为(如果赛区设置报名号的话):							
所属学校(请填写完整的全名):福州大学							
参赛队员 (打印并签名): 1. 李译 (13514076602)							
2 李志坤 (13599393672)							
3殷婷							
指导教师或指导教师组负责人 (打印并签名):							
日期: 2007 年 9 月 24 日							

赛区评阅编号(由赛区组委会评阅前进行编号):

# 2007 高教社杯全国大学生数学建模竞赛

# 编号专用页

赛区评阅编号(由赛区组委会评阅前进行编号):

赛区评阅记录(可供赛区评阅时使用):

	 7 / 1 / 2 / C				
评阅人					
评分					
备注					

全国统一编号(由赛区组委会送交全国前编号):

全国评阅编号(由全国组委会评阅前进行编号):

### 中国人口增长预测

#### 摘要:

针对题目所提要求,我们建立了两个中国人口预测模型,分别用于对中国人口的发展趋势做短期和中长期的预测。

为了对中国人口发展做短期的预测,考虑到题目所给的数据资料的不全面,我们由马尔萨斯的人口指数增长模型得到启发,针对中国人口发展的特点,把出生率和死亡率函数这两大对人口增长起主要作用的因素作为建模的关键参数,在附件中没有给出中国近年总人口数的情况下,建立了短期内预测中国人口增长的微分方程模型。在该模型中,为了得到出生率和死亡率函数这两个重要参数,我们通过分析题目所给数据,提取出有效信息,计算归纳出 2001 年到 2005 年的出生率和死亡率,并在此基础上引入灰色模型,用于对出生率和死亡率进行预测,得出了出生率和死亡率关于时间的函数。较准确的估计出了人口增长的关键参数,使得建立的人口增长短期预测模型不仅符合中国人口的发展特点,而且简单易用,能在未知总人口数的情况下预测人口的相对发展变化,这一优点使得可以方便且准确的用于预测中国人口短期内的发展趋势。

为了对中国人口发展做中长期的预测,考虑到短期模型在预测人口中长期发展中的局限性以及影响人口发展的众多因素的不确定性和它们之间关系的复杂性,我们利用灰色动态模型的特点,从《中国统计年鉴》中查到了中国近年的人口总数(见附表一),把人口数做为灰色量,对原始各年人口序列进行分段建模,对各分段模型进行定性分析比较,根据各阶段宏观指标的相关确定一组适当的权数,进行预测模型的最优组合,以确定最优预测模型,从而建立了中长期预测中国人口增长的灰色动态系统人口模型,对中国人口进行了中长期的预测。

在对中国总人口进行短期和中长期的总体预测后,我们从附件中提取出城、镇、乡三地人口、男女出生性别比、妇女生育率、老龄人口比率等相关数据,对中国未来城、镇、乡三地人口比例、男女出生性别比、妇女生育率、老龄人口比率等影响人口发展的主要因素做趋势预测,从而达到了对中国人口全方位的预测。

**关键词:** 出生率、死亡率、指数增长模型、灰色动态模型、性别比、老龄化、生育率。

### 一、问题重述

中国是一个人口大国,人口问题始终是制约我国发展的关键因素之一。根据已有数据,运用数学建模的方法,对中国人口做出分析和预测是一个重要问题。

近年来中国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。2007年初发布的《国家人口发展战略研究报告》(附录 1)还做出了进一步的分析。

关于中国人口问题已有多方面的研究,并积累了大量数据资料。附录 2 就是从《中国人口统计年鉴》上收集到的部分数据。

试从中国的实际情况和人口增长的上述特点出发,参考附录2中的相关数据(也可以搜索相关文献和补充新的数据),建立中国人口增长的数学模型,并由此对中国人口增长的中短期和长期趋势做出预测;特别要指出你们模型中的优点与不足之处。

## 二、基本假设

- 1、把市镇乡看成独立封闭的系统,即不考虑迁入迁出的因素对人口的影响。
- 2、妇女的总和生育率设为1.8。
- 3、忽略经济,社会环境,资源等因素对人口的影响。

#### 三、符号说明

City(T): 附件所给数据中的第 T 年城市人数

Town (T): 附件所给数据中的第 T 年城镇人数

Village(T): 附件所给数据中的第 T 年的乡村人数

 $M_i$ : i=1,2,3 分别表示城市,城镇,乡村男性比率

 $W_i$ : i=1,2,3 分别表示城市,城镇,乡村女性比率

 $Md_i$ : i=1,2,3 分别表示城市,城镇,乡村男性死亡比率

Wd: i=1,2,3 分别表示城市,城镇,乡村女性死亡比率

Deads: 死亡人数 Borns:出生人数

 $B_i$ : 育龄妇女比率(15 岁~49 岁) i=1,2,3 分别表示城市,城镇,乡村育龄妇女比率

 $Bc_i$ : i=1,2,3 分别表示城市,城镇,乡村女性生育率

T: 年份

t:年份表达变量

b(t):出生率随时间变化的函数

d(t):死亡率随时间变化的函数

N(t):总人口数随时间变化的函数

No: 2001 年人口数量(见附录一)

## 四、问题分析与建模及人口预测

中国人口增长的数学模型:

4.1 用于短期预测的模型:

根据附件所给材料中的数据,计算第 T 年的死亡人数及出生人数:公式为:

Deads=City(T)\*(M1\*Md1+W1\*Wd1)+Town(T)\*(M2\*Md2+W2\*Wd2)+Village(T)\*(M3\*Md3+W3\*Wd3)

Borns=City(T)\*B1\*Bc1+Town(T)\*B2\*Bc2+Village(T)\*B3\*Bc3

计算第 T 年的出生率 b(t)=Borns/(City(T)+Village(T)+Town(T))

得如下表格:

T	t	出生率 b(t)	死亡率 d(t)
2001	0	0. 01740845	0.00963373
2002	1	0. 016390372	0.00936573
2003	2	0. 015774741	0.00911349
2004	3	0. 016341515	0.00896989
2005	4	0. 014339363	0.00789073

对于序列  $X(t)=\{X(0),X(1),X(2)\cdots X(n)\}$  根据灰色模型中较常用的 GM(1,1) 模型:

$$\frac{dX(t)}{dt} + aX(t) = u \tag{1}$$

来预测 b(t), d(t) 的趋势。

(1) 参数 a, u 的估计及 X(t) 预测方程的求解

将 (1) 写成  $a[-X(t)] + u = \frac{dX(t)}{dt}$ , 将 t 换为 t+1 并与原式做算术平均得:

$$a\{-\frac{1}{2}[X(t) + X(t+1)]\} + u = \frac{1}{2}[X'(t) + X'(t+1)]$$
 (2)

我们可以用差分DX(t)=X(t+1)-X(t)近似代替(2)式右端得:

$$a\{-\frac{1}{2}[X(t) + X(t+1)]\} + u = DX(t)$$

记  $A=\begin{pmatrix} a \\ u \end{pmatrix}$ ,用最小二乘法估计出系数矩阵 A。 $A=\begin{pmatrix} B^TB \end{pmatrix}^{-1}\begin{pmatrix} B^TC \end{pmatrix}$ 。其中,

$$B = \begin{bmatrix} -\frac{1}{2}(X(0) + X(1)) & 1\\ -\frac{1}{2}(X(1) + X(2)) & 1\\ \dots & \dots\\ -\frac{1}{2}(X(n-1) + X(n)) & 1 \end{bmatrix}$$

$$C = (DX(0), DX(1)...DX(n-1))^{T}, (DX(t) = X(t+1) - X(t))$$

有了 a,u 的估计值之后,我们就可以求解(1)的微分方程。

(1) 式两端同乘以
$$e^{at}$$
得 $e^{at}\frac{dX(t)}{dt}+ae^{at}X(t)=e^{at}u$ ,

可化为
$$\frac{d}{dx}$$
{ $e^{at}X(t)$ } =  $e^{at}u$  (3)

两边取不定积分得 $e^{at}X(t) = \int e^{at}udt + c$ (c 为待定常数)

解得  $X(t)=ce^{-at}+\frac{u}{a}$ ,将 t=0 代入得  $c=X(0)-\frac{u}{a}$ 。所以有 X(t)预测方程:

$$X(t)=[X(0)-\frac{u}{a}]e^{-at}+\frac{u}{a}$$

#### (2) 短期预测模型

将 b(t),d(t) (t=0~4)的值代入上述 GM(1,1)模型的 X(t)中求解。

<1>対于 b(t)={ 0.01740845, 0.016390372, 0.015774741, 0.016341515, 0.014339363}

得 
$$B_b = \begin{pmatrix} -0.02027925 & 1 \\ -0.0192977 & 1 \\ -0.01926975 & 1 \\ -0.0184084 & 1 \end{pmatrix}$$

 $C_b = (-0.001018078, -0.000615631, 0.000566774, -0.002002152)^T$ 

$$A_b = (B_b^T B_b)^{-1} (B_b^T C_b) = (-0.21878, -0.00515288)^T$$

a=-0.21878,u=-0.00515288

得预测方程:

$$b(t) = -0.00266266e^{0.21878t} + 0.0235528 \tag{4}$$

<2> 对于 d(t)={ 0.00963373, 0.00936573, 0.00911349, 0.00896989, 0.00789073}

$$B_d = \begin{pmatrix} -0.00949973 & 1 \\ -0.00923961 & 1 \\ -0.00904169 & 1 \\ -0.00843031 & 1 \end{pmatrix}$$

 $C_d = (-0.000268, -0.00025224, -0.0001436, -0.00107916)^T$ 

$$A_d = (B_d^T B_d^T)^{-1} (B_d^T C_d^T) = (-0.209686, -0.0023334)^T$$

得预测方程

$$d(t) = -0.00149437e^{0.209686t} + 0.0111281$$
 (5)

所以增长率函数为  $g(t) = 0.00149437e^{0.209686t} - 0.00266266e^{0.21878t} + 0.0124247$ 则对总人口数  $\mathbf{N}(t)$ ,有:

$$\frac{dN(t)}{dt} = g(t)N = (0.00149437e^{0.209686t} - 0.00266266e^{0.21878t} + 0.0124247)N$$
(6)

且有初值条件 N(0)=  $N_0$ =12.7627 亿人( $N_0$ 为 2001 年的总人口数,由参考文献[]得)

解该微分方程(6)得:

$$N(t) = N_0 e^{0.0071267 e^{0.209686t} - 0.01217 e^{0.21878t} + 0.0124247t + 0.0050433}$$

此式可用来预测短期内(附件所给数据截至 2005 年, 预测期为 2006~2011 年)的人口趋势

T	t	总人口数(亿人)
2006	5	13.4327
2007	6	13.5444
2008	7	13.6424
2009	8	13.7228
2010	9	13.7805
2011	10	13.8091

#### 4.2 用于中长期预测的模型:

鉴于影响人口发展的众多不确定因素,我们应用灰色动态模型 GM(1,1). 为了得到最准确的预测结果,在选取数据年份和年数上,我们进行了认真的分析 验证。数据筛选过程如下: 1、我们选取了1970—1980 和1980—1990这两段时间的人口数,带入灰色动态模型中,计算出1995-1999这个时间段的人口数,和查得的资料数据进行比较,得到表格如下:

年	实际总人口 (万人)	1970-1980 预 测值	预测误 差	1980-1990 预测值	预测误 差	1965-1980 预测值	预测误差
1995	121121	120866	0.21	122651	1.26	129429	6.86
1996	122389	122345	0.04	124330	1.59	131458	7.41
1997	123626	123825	0.16	126004	1.92	133622	8.09
1998	124810	125309	0.4	127694	2.31	135870	8.86
1999	125909	126901	0.71	129404	2.78	138151	9.72

结论1:年份越近,预测越准。

2、我们选取1989-1994 6年和 1988-1995 8年的时间段的人口数,带入灰色动态模型中,计算出1996-1999这个时间段的人口数,和查得的资料数据进行比较,得到表格如下:

年		实际总人口(万	1989-1994 预测	预测误	1988-1995 预	预测误
		人)	值	差	测值	差
	1996	122389	122216	0.141	122543	0.126
	1997	123626	123337	0.234	123796	0.138
	1998	124810	124400	0.329	125052	0.194
	1999	125909	126066	0.325	126309	0.318

结论2: 年份越少, 预测越准。

于是,我们选取1999-2004这六年的人口数据,带入到灰色动态模型中,得到灰色动态人口模型:

N(t)=17.6260714-17.73150933 e<sup>-0.0225527t</sup> (规定1999年时t=0)

并用该模型预测人口数如下:

年	2007	2008	2009	2010	2015	2020	2025
预测人口 数	133028	133734	134404	135042	137780	139888	141496
年	2030	2035	2040	2045	2050		
预测人口 数	142736	143696	144464	145104	145600		

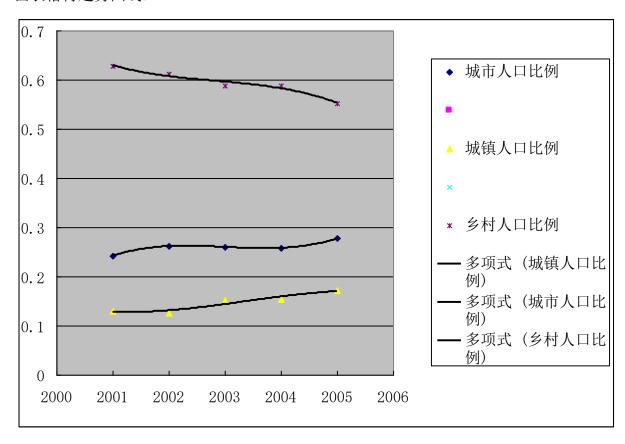
## 4.3 人口迁移对人口增长的影响

由附件所给数据求得  $2001^22005$  年城市,城镇,乡村人口分别所占总人口的比例 K(T)= 某 区 域 ( city or Town or Village ) 人 口 数 据 / 总 人 口 数 据 (City(T)+Town(T)+Village(T)) 得如下表格:

年份 | 城市人口比例 | 城镇人口比例 | 乡村人口比例 |

2001	0. 242	0. 1297	0. 6283
2002	0. 2616	0. 1255	0. 6129
2003	0. 2602	0. 1522	0. 5876
2004	0. 2582	0. 1536	0. 5882
2005	0. 2772	0. 1713	0. 5516

## 由表格得趋势图线:

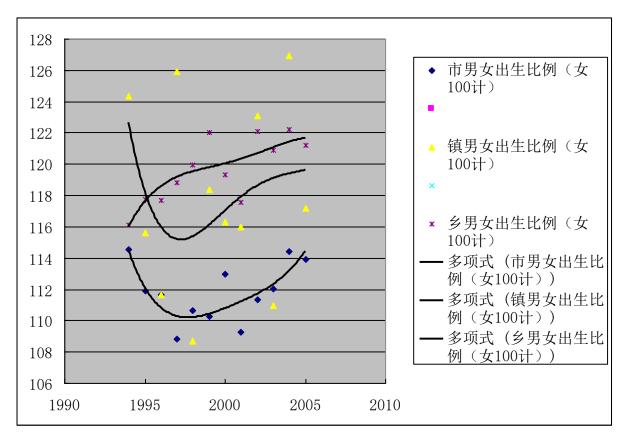


由趋势可见,城市与城镇的人口比例在上升,而乡村的人口比例在减少,可见如今乡村人口正在逐步往城市和城镇迁移。这对于我国未来人口的发展不能说没有影响,因为乡村人口由于计划生育等措施的普及没有城市与城镇广,导致其生育率相对较高,而且乡村人口基数相对较大,所以对过去的人口增长有很大影响。现在随着乡村人口往城市与城镇慢慢迁移,将使得总体生育率有所下降,这就能很有效的控制人口的增长。

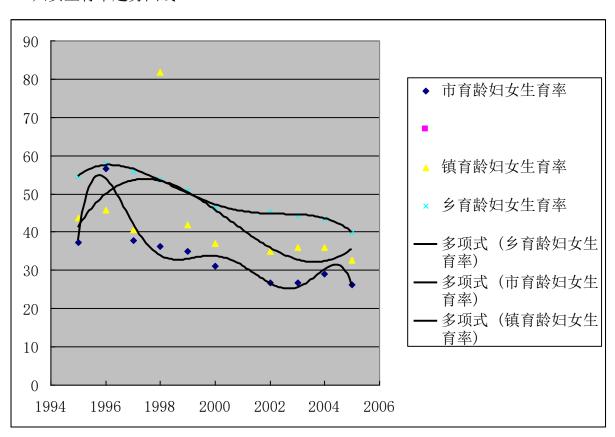
#### 4.4 性别比及生育率对人口数的影响

由附件所给1994~2005年的性别比数据和妇女生育率数据用Excel 绘制趋势图得:

## <1>性别比数据图线



## <2>妇女生育率趋势图线



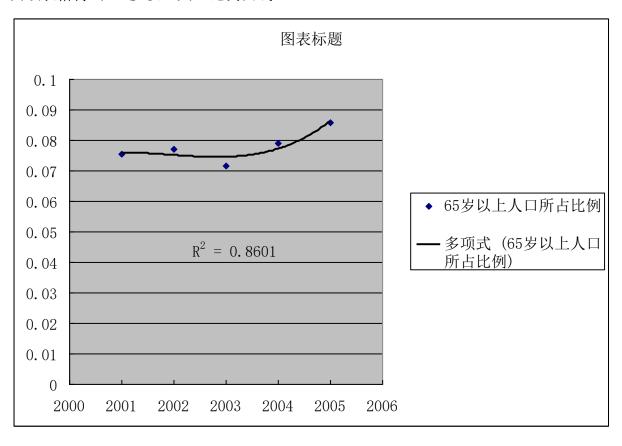
由图象可看出:自1997年后,男性出生比例总体处于上升趋势,尤其是在占人口比例较大的乡镇。女性出生比例的相对下降,与此同时,妇女的生育率却在1998年左右开始缓慢下降,这无疑对中国未来总人口的发展有比较大的影响。这些数据表明在未来一定时期内人口的增长率将逐年降低,随着时间的推移,相应政策的实施及高生育率的乡镇人口往城市迁移,预计未来几十年后,人口将趋于零增长甚至负增长。

## 4. 5人口老龄化对我国人口的影响

由附件所给数据得出  $2001^22002$  年 65 岁以上人口占总人口的比例, 计算公式为: K(T)=65 岁以上人口数据/总人口数据 (City(T)+Town(T)+Village(T)), 得如下表格:

年份	65 岁以上人口所占比例
2001	0.07536
2002	0. 07696
2003	0.07169
2004	0. 07897
2005	0. 08571

由该表格得出65岁以上人口比例曲线:



可见老龄人口数量总体上在逐年增多,老龄化进程加快。资料表明,自 1999年中国进入老龄社会开始,老年人口数撤不断增加,老龄化程度持续加深。老龄化的加深势必导致总人口的上升。中国将面临人口老龄化和人口总量过多的

双重压力。目前人口总规模高达 13 亿,预计到 2030 年达到最大人口规模 14.65 亿,总人口。过多的压力将长期存在。与此同时,人口老龄化压力已经开始显现。并将随着老龄化的发展而不断加重。整个 21 世纪,这两方面压力将始终交织在一起,给中国经济、社会发展带来严峻的挑战。

### 五、模型优缺点分析及改进方向

人口模型是很古老的研究课题,前人已有很多成熟的模型,侧重各个方向, 所以适用范围有所不同,因此,我们在前人已建立的模型的基础上,建立了适合 中国人口发展趋势的模型。

- 5.1 模型的优点:
- 5.1.1 微分方程模型的优点:

该模型是在马尔萨斯的人口指数增长模型的基础上,增加了出生率和死亡率 这两大影响人口变化的关键因素,在未知总人口数的情况下,预测了人口的相对 发展趋势并为简化计算,忽略了其他因素对人口变化的影响,使得所建立的模型 不仅承传了指数增长模型简单的优点,而且能比其更准确的预测出中国人口的增 长。是一个较实用的中国人口短期预测模型。

5.1.2 灰色系统人口模型的优点:

对于影响人口系统的因素,出了出生率和死亡率外,还有净迁入量,社会经济,自然环境,科学技术等一系列方面,这些众多的因素,不是用几个指标所能表达清楚的.而且,它们之间的结构关系错综复杂,它们对人口增长的作用更是无法精确计算.多数因素都在动态变化之中,其运行机制和变化规律难以完全明白.所以,将灰色模型用到人口预测中不仅简单而且能达到比较准确的预测效果。

- 5.2 模型的缺点:
- 5.2.1 微分方程模型的缺点:
- 1、未考虑人口迁移率对人口的影响,影响预测的准确度;
- 2、未考虑出生人口性别比的不断增大对人口发展的影响;
- 4、对死亡率的计算中,一般化了婴儿的死亡率,即没有把婴儿的死亡率单独出来计算,导致对死亡率的预测不准;
- 5、没有分年龄、男女对人口进行预测,使得预测不够详细具体。

#### 5.2.2 灰色预测模型的缺点:

- 1、GM(1,1)模型在其使用条件上存在着一定的限制,它是描述按指数规律变化的事物的模型,因此,它使用于呈指数规律发展变化的系统进行预测。也就是说我们默认了人口的指数增长方式。
- 2、在1的前提下,人口的发展速度却可能由于各种因素的影响而出现阶段性增长的特征,呈"波浪"型发展,那么利用 GM(1,1)模型进行预测时,由于累加生成数列的缘故,也会对其发展的规律产生影响,削弱人口系统发展的阶段性规律。

#### 5.3 模型的改进:

1、可以把年龄作为变量建立非线性模型:

$$X_{i+1}(t+1) = S_i(t)X_i(t)$$
  $i = 0,1,2,\dots,92.$ 

$$X_0(t) = \beta(t)\sum_{18 \leq i \leq 49} K_i(t)H_i(t)X_i(t)$$

其中 $X_i(t)$ 为 t 年 i 岁人口总数,  $S_i(t)$ 为 t 年 i 岁人口存活率,  $K_i(t)$ 为 t 年 i

岁人口妇女所占比例, **H(t)**为 t 年分年龄生育模式函数。

2、把男性和女性分开,分别建立男性和女性的模型:

$$\frac{dM(t)}{dt} = b(t) * \sqrt{M(t) * F(t)} - d(t) * M(t)$$

$$\frac{dF(t)}{dt} = b(t) * \sqrt{M(t) * F(t)} - d(t) * F(t)$$

- M(0)和F(0)为初始人口数,为已知。
- M(t)为女性人口,F(t)为男性人口

## 六、参考文献

## [1]未知作者 中国数据在线

http://www.chinadataonline.org/member/yearbook/default.asp?StartYear=1984&EndYear=2005> 2007.9.21.

- [2] 王宏健 全国大学生数学建模竞赛优秀论文汇编 福建 2007
- [3] 谢剑,赵彤 <<Excel 在建筑工程中的应用:快速计算工具实例 < http://search.dangdang.com/rd.asp?id=8792892&clsid=01.54.11.00&key=%ca%fd%be%dd%c4%e2%ba%cf>>> 天津大学出版社 2004
- [4]邓聚龙 灰色系统理论教程[M]. 华中理工大学出版社, 19901
- [5] 姜启源 数学模型[M] 北京: 高等教育出版社 2005
- [6] 同济大学应用数学系 工程数学线性代数 高等教育出版社 2005
- [7] 中华人民共和国国家统计局. 中国统计年鉴-2002[M]. 北京: 中国统计出版社, 2002
- [8] 中华人民共和国国家统计局, 中国统计年鉴-2001[M], 北京: 中国统计出版

#### 社,2001

[9] 中华人民共和国国家统计局. 中华人民共和国 2002 年国民经济和社会发展统计公报[J]. 中国统计, 2003(3)

[10] 刘思峰, 郭天榜, 党耀国. 灰色系统理论及其应用[M]. 北京: 科学出版社, 1999

[11]于学军 对第五次全国人口普查数据中总量和结构的估计 人口研究 2002.3 [12]倪少凯、陈卫红 用实现灰色数列模型 GM(1,1)的预测 树理医药学杂志 2002, 15(2)292~293

[编号] 作者,论文名,杂志名,卷期号:起止页码,出版年。

#### 附表一:

年份 年末总人口 出生率 死亡率 自然增长率 总和生育率 (**%**<sub>0</sub>) (%) (%) (万人) (个) 1949 54167 36.00 20.00 16.00 6.14 1950 55196 37.00 18.00 19.00 5.81 1951 56300 37.80 17.80 20.00 5.70 1952 57482 37.99 17.00 20.99 6.47 1953 37.00 58796 14.00 23.00 6.05 1954 60266 38.19 6.28 13.18 25.00 1955 61465 32.18 12.28 19.90 6.26 1956 62780 33.67 11.40 21.39 5.85 1957 64238 34.03 23.23 10.80 6.41 1958 65346 29.22 11.98 17.24 5.68 1959 66012 24.78 14.59 10.19 4.30 1960 66207 20.86 17.91 2.95 4.02 1961 66457 18.02 14.24 3.78 3.29 1962 67295 22.63 12.61 10.02 6.02 7.50 1963 69172 40.00 12.11 27.89 1964 70499 30.68 11.50 19.18 6.18 1965 72538 38.42 9.50 28.92 6.08 1966 74206 31.82 22.99 8.83 6.26 1967 76032 33.04 8.43 24.61 5.31 1968 78198 36.70 8.21 28.49 6.45 1969 80335 35.35 8.03 27.32 5.72 1970 82542 35.07 7.60 27.47 5.81 27.10 1971 84779 34.42 7.32 5.44 1972 86727 30.59 7.61 22.98 4.98 1973 88761 30.49 7.04 23.45 4.54 1974 90409 25.91 7.34 18.57 4.17 1975 17.27 91970 24.59 7.32 3.57 1976 21.35. 7.25 14.10 93267 3.24 1977 94774 23.03 6.87 16.16 2.84 1978 96159 20.86 6.25 14.61 2.72

1979	97542	20.59	6.21	14.38	2.75
1980	98705	18.26	6.34	11.92	2.24
1981	100072	20.21	6.36	13.85	2.63
1982	101654	22.28	6.60	15.68	2.87
1983	103008	20.19	6.90	13.29	2.42
1984	104357	19.90	6.82	13.08	2.35
1985	105851	21.04	6.78	14.26	2.20
1986	107507	22.43	6.86	15.57	2.42
1987	109300	23.33	6.72	16.61	2.59
1988	111026	22.37	6.64	15.73	2.31
1989	112704	21.58	6.54	15.04	2.25
1990	114333	21.06	6.67	14.39	2.17
1991	115823	19.68	6.70	12.98	2.01
1992	117171	18.24	6.64	11.60	
1993	118517	18.09	6.64	11.45	
1994	119850	17.70	6.49	11.21	
1995	121121	17.12	6.57	10.55	
1996	122389	16.98	6.56	10.42	
1997	123626	16.57	6.51	10.06	
1998	124761	15.64	6.50	9.14	
1999	125786	14.64	6.46	8.18	
2000	126743	14.03	6.45	7.58	1.22*
2001	127627	13.38	6.43	6.95	
2002	128453	12.86	6.41	6.45	
2003	129227	12.41	6.40	6.01	
2004	130000				