

基于差分方程的人口预测模型

摘要

本文针对我国人口增长中出现的新特点，建立了两个符合实际情况的预测模型。

模型一：基于 Logistic 模型，建立了含市、镇、乡人口相互流动关系的微分方程模型，求得全国总人口数在短期内将持续增长，到 2010 年、2020 年分别为 13.59 亿和 14.44 亿，具有较好的中短期预测效果。

模型二：从年龄转移与总和生育率出发，建立了离散型人口发展模型。

1. 针对性别比例，引入女性比例转移矩阵，利用计算机进行随机模拟，建立起动态的女性比例转移矩阵；
2. 针对人口迁移，引入人口迁移率矩阵，将迁移率标准化后利用平均迁移率实现了对迁移人数的预测；
3. 针对死亡率，利用分段加权法估计其随时间的变化，得到了较好的预测结果；
4. 针对老龄化和出生高峰，将其转化为育龄妇女占总人口的比例，实现了量化预测。

综合考虑上述因素，利用 MATLAB 编程求解，全国总人口将呈现短期增加、中期平稳、长期缓慢下降的趋势。其中总人口在 2010 年、2020 年将分别达到 13.63 亿和 14.57 亿人；在 2039 年将达到人口高峰 14.65 亿人，性别比逐步逼近 46.7%，69 岁以上人口将超过 4 亿人，占总人口的 27.48%。

模型一需要的原始数据少，操作简单，适合于中短期预测，但长期预测效果不佳；模型二综合考虑了各因素，对中短期和长期均有较好的预测效果，但所需数据量大，操作较为复杂。

关键字 Logistic 模型 人口发展模型 转移矩阵 计算机模拟 迁移率

1. 问题重述

人口预测是国家工作中的重点，关系着国家的发展方向和命运。我国是一个人口大国，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，对我国人口做出分析和预测是一个重要问题。

近年来我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，这些都影响着我国人口的增长。2007 年初发布的《国家人口发展战略研究报告》对此做出了进一步的分析。

从我国的实际情况和人口增长的上述特点出发，参考相关数据，建立我国人口增长的数学模型，并由此对我国人口增长的中短期和长期趋势做出预测；特别注意指出模型中的优点与不足之处。

2. 问题分析

一个社会（国家、省市、地区）人口的变化和随时间的发展过程，是由很多因素决定的，社会制度、自然环境、生活水平、科学文化水平、战争、自然灾害和移民等等，都能严重地影响社会人口的发展过程。然而，婴儿的出生、人口的死亡、居民的迁移却是决定该社会人口变化的直接原因，近年来我国人口发展出现的一些新特点，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，都直接或间接地通过这三个现象表现出来。综合考虑这些因素成为构建符合我国国情的人口增长模型关键。

建立模型对人口发展过程进行定量预测，就是根据现有的人口统计资料和原始数据，从当前实际的人口状况出发，并对未来的人口发展过程，提出合理的控制要求和假定，应用科学的方法，预测出未来几年、几十年甚至上百年的人口发展趋势，包括人口总数、人口的性别、年龄和城乡结构，人口出生、死亡和自然增长率的变化以及在未来的人口构成中劳力和抚养水平及老化水平等。

3. 模型假设

针对本问题，建立如下合理的假设：

1. 题中所给数据能反映我国人口变化的基本情况；
2. 一些重大事件，如战争、自然灾害等对人口预测的影响暂不考虑；
3. 所给数据都是年末数据，也即下一年年初数据，如 2001 年总人口实质上也表示 2002 年初的总人口；
4. 今年所统计的 i 岁的人口在下一年年初均为 $i+1$ 岁；
5. 生育模式不随时间变化。

4. 符号说明

| 符号 | 说明 | 单位 |
|------------|---------------------------|----|
| $N(t)$ | 在 t 年初的总人口数 | 万人 |
| N_m | 最大人口容量 | 万人 |
| r | 人口固有增长率 | \ |
| l | 各地区间人口迁移系数 | \ |
| $x_i(t)$ | 第 t 年年初有 i 岁的人口数 | 万人 |
| $m_i(t)$ | 第 t 年 i 岁人的死亡率 | \ |
| $g_i(t)$ | 第 t 年 i 岁人口的迁入量 | 万人 |
| $k_i(t)$ | 第 t 年女性人口在 i 岁人口中所占比例 | \ |
| $h_i(t)$ | 第 t 年生育模式 | \ |
| $b(t)$ | 第 t 年总和生育率 | \ |
| $[a1, a2]$ | 育龄期范围, $a1 = 15, a2 = 49$ | 岁 |
| $h_i(t)$ | 第 t 年 i 岁人的后向死亡率 | \ |
| $m_i(t)$ | 第 t 年 i 岁人的前向死亡率 | \ |
| $s_i(t)$ | 第 t 年 i 岁人的迁移率 | \ |

5. 模型建立与求解

5.1. 数据预处理

题中所给 5 年我国人口 1% 调查数据是对人口的抽样调查数据, 由于数据的不完备性, 并不能由它来估计当时的全国总人口数。但基于抽样调查的等概率性, 可以认为它所反应的市、镇、乡三个地区的人口比例及男女比例是与实际较为接近的。从《中国人口统计年鉴 2006》^[1]可以得到 2001~2005 年具体的全国总人口数。进而可以得到各部分人口数。所得数据见表 1。

例: 2001 年市实际男性总人口 = 2001 年全国总人口 $\times \frac{\text{当年抽样调查中市男性总人口}}{\text{当年抽样调查总人口}}$

表 1 5 年间各部分人口数

| 年份 | 2001 | 2002 | 2003 | 2004 | 2005 |
|-----|--------|--------|--------|--------|--------|
| 城市男 | 15466 | 16803 | 16779 | 16677 | 18149 |
| 城市女 | 15420 | 16804 | 16843 | 16889 | 18092 |
| 镇男 | 8394 | 8213 | 9961 | 10060 | 11276 |
| 镇女 | 8154 | 7904 | 9705 | 9901 | 11117 |
| 乡男 | 41271 | 40566 | 39222 | 39360 | 36661 |
| 乡女 | 38923 | 38164 | 36718 | 37100 | 35461 |
| 总人口 | 127627 | 128453 | 129227 | 129988 | 130756 |

5.2. 模型一：基于人口迁移的 Logistic 阻滞增长模型

5.2.1. 模型建立：

考虑单一地区只受自然资源和环境条件限制的 Logistic 阻滞增长模型^[2]：

$$\frac{dN(t)}{dt} = rN(t)\left(1 - \frac{N(t)}{N_m}\right)$$

其中 r 为固有增长率。 $N(t)$, N_m 分别表示第 t 年的人口数和此条件限制下的最大人口容量。因子 $rN(t)$ 体现了人口自身的增长趋势, 因子 $(1 - N(t)/N_m)$ 则体现了资源和环境对人口增长的阻滞作用, 人口增长是两个因子共同作用的结果。

现在考虑市、镇、乡三个地区的总人口变化规律, 假定在一段时期内, 这三个地区各自均处于稳定状态, 出生率和死亡率都没有太大变化。然而居民在三个地区间的迁移会对总人口产生一定的影响。对原 Logistic 模型进行修改, 得到多地区条件下的改进模型：

$$\begin{cases} \frac{dN_1(t)}{dt} = r_1 N_1(t) \left(1 - \frac{N_1(t)}{N_{1m}} + I_{21} \frac{N_2(t)}{N_{2m}} + I_{31} \frac{N_3(t)}{N_{3m}}\right) \\ \frac{dN_2(t)}{dt} = r_2 N_2(t) \left(1 - \frac{N_2(t)}{N_{2m}} + I_{12} \frac{N_1(t)}{N_{1m}} + I_{32} \frac{N_3(t)}{N_{3m}}\right) \\ \frac{dN_3(t)}{dt} = r_3 N_3(t) \left(1 - \frac{N_3(t)}{N_{3m}} + I_{13} \frac{N_1(t)}{N_{1m}} + I_{23} \frac{N_2(t)}{N_{2m}}\right) \end{cases}$$

其中 $p, q = 1, 2, 3$ 分别表示市、镇、乡这三个地区, I_{qp} 是迁移系数, 正数时表示迁入, 负数时表示迁出。 I_{qp} / N_m 体现了居民从地区 q 迁到地区 p 时对地区 p 人口增长的影响。

为求解此方程, 不失一般性的取 $\Delta t = 1$, 则原微分方程组化为差分方程组：

$$\begin{cases} N_1(t+1) - N_1(t) = r_1 N_1(t) \left(1 - \frac{N_1(t)}{N_{1m}} + I_{21} \frac{N_2(t)}{N_{2m}} + I_{31} \frac{N_3(t)}{N_{3m}}\right) \\ N_2(t+1) - N_2(t) = r_2 N_2(t) \left(1 - \frac{N_2(t)}{N_{2m}} + I_{12} \frac{N_1(t)}{N_{1m}} + I_{32} \frac{N_3(t)}{N_{3m}}\right) \\ N_3(t+1) - N_3(t) = r_3 N_3(t) \left(1 - \frac{N_3(t)}{N_{3m}} + I_{13} \frac{N_1(t)}{N_{1m}} + I_{23} \frac{N_2(t)}{N_{2m}}\right) \end{cases}$$

5.2.2. 模型求解：

1. 求解 r

考虑到育龄妇女生育率信息的残缺, 用年龄为 i 的育龄期妇女占总人口的比例 $p_i(t)$ 与对应每个年龄生育率的乘积和来表示年 t 的出生率。

$$r = \sum_{i=a1}^{a2} p_i(t) b_i(t)$$

取 $t = 2001$, 可得到市、镇、乡的固有增长率： $r_1 = 0.00907$, $r_2 = 0.0107$, $r_3 = 0.01233$

2. 求解 N_m

由经验公式 $N_m = N(t)/(1-r/r_0)$,其中 $r_0 = 0.029$,为世界人口固有增长率 ,可得 2001 年各地区最大人口容量 (单位 : 万人) : $N_1 = 139488$, $N_2 = 26228$, $N_3 = 44944$ 。

3. 求解 I

为求解 I ,把 2002 年和 2001 年的数据都带入阻滞增长模型 ,联合以下三个迁入迁出方程可解迁移系数 I 。(见表 2)

$$\begin{cases} \frac{r_1 I_{21}}{N_{2m}} = \frac{r_2 I_{12}}{N_{1m}} \\ \frac{r_2 I_{32}}{N_{3m}} = \frac{r_3 I_{23}}{N_{2m}} \\ \frac{r_3 I_{13}}{N_{1m}} = \frac{r_1 I_{31}}{N_{3m}} \end{cases}$$

表 2 迁移系数 I

| | 市 $p=1$ | 镇 $p=2$ | 乡 $p=3$ |
|---------|------------|----------|------------|
| 市 $q=1$ | 0 | 21.13720 | -0.1108025 |
| 镇 $q=2$ | 14.55182 | 0 | 3.735552 |
| 乡 $q=3$ | -0.4674880 | 22.89316 | 0 |

4. 求解 $N_p(t+1)$

由所求得的系数 r, N_m, I ,把 2001 年作为初始年 ,用 MATLAB 求解可得以后各年全国总人口数 :

表 3 模型一对全国总人口数的预测

| 年代 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 |
|-------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 预测人数 (亿人) | 12.94 | 13.03 | 13.12 | 13.22 | 13.32 | 13.41 | 13.50 | 13.59 | 13.68 | 13.77 |
| 实际人数 (亿人) | 12.92 | 13.00 | 13.08 | —— | —— | —— | —— | —— | —— | —— |
| 误差 (%) | 0.13 | 0.26 | 0.39 | —— | —— | —— | —— | —— | —— | —— |
| 年代 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 |
| 预测人数 (亿人) | 13.86 | 13.94 | 14.03 | 14.11 | 14.19 | 14.28 | 14.26 | 14.44 | 14.52 | 14.61 |
| 年代 | 2023 | 2024 | 2025 | 2026 | 2027 | 2028 | 2030 | 2035 | 2040 | 2045 |
| 预测人数 (亿人) | 14.69 | 14.78 | 14.86 | 14.95 | 15.04 | 15.12 | 15.30 | 15.76 | 16.24 | 16.82 |

5.2.3. 模型分析 :

1. 结果分析

由表 3 可以看出 ,我国人口在短时间内仍会继续增长 ,增长水平为每年 1000 万人左右 ,到 2010 年和 2020 年 ,全国总人口将达 13.59 亿人与 14.44 亿人。

2. 敏感性分析

由模型的建立原理可知： $N(t) \propto r(1-r/N_m) \propto r(1-r)$ ，故 $N(t)$ 将随 r 的增加成二次幂增长，而 $r = \sum \text{女性比例} \times \text{生育模式}$ ，女性比例变化不大的情况下，可认为 r 正比于生育模式，又因为生育模式正比于总和生育率，故 $N(t)$ 与总和生育率成二次正相关，即整个模型对总和出生率十分敏感。

3. 模型优点

基于传统的 Logistic 模型，建立起市、镇、乡人口间的相互流动关系，综合考虑了出生率、死亡率与迁入迁出对人口增长的影响，对 Logistic 原模型做了合理的扩充，使其更符合实际情况，预测效果良好。同时根据实际情况，对难以求解的微分方程采用差分方程代替，降低了解题难度，使模型易于实现。

4. 模型缺点

- i. 由于采用了差分方程的近似代替，整个预测结果均随着时间的增长而增大，这是不符合生物学规律的，故此模型对人口的中短期预测较好，长期预测则难以让人信服。
- ii. 没有考虑性别比对年龄结构的影响，且只能对总人口数进行预测，无法对性别比和年龄结构做出相应的预测。

5. 改进方向：

- i. 采用先进的工具软件对微分方程求解，得到模型预测的准确结果。
- ii. 分段对模型进行调整和改进，减少累加误差。

5.3. 模型二：离散人口发展方程模型

5.3.1. 模型建立分析

由于模型一的局限性，需要进一步分析各因素对全国人口总数的影响。考虑一个地区第 $t+1$ 年初的人口，它有三类主要来源：

- i. 本地区第 t 年存活下来的人；
- ii. 本地区第 t 年出生并存活下来的人；
- iii. 外部地区第 t 年迁入或迁出本地区的人。

为定量描述这些因素对人口数的影响，建立如下人口发展模型^[3]：

5.3.2. 模型建立

针对一个地区 p ，若第 t 年年初有 i 岁人口数 $x_i(t)$ 人，则由假设 4 第 $(t+1)$ 年年初这些人为 $(i+1)$ 岁。若 $m_i(t)$ 为这批人在一年内的死亡率， $g_i(t)$ 为迁移进来的人口数，则 $(t+1)$ 年年初 $(i+1)$ 岁的人口数为 $x_i(t)(1-m_i(t)) + g_i(t)$ 。0 岁人口数 $x_0(t)$ 需通过妇女生育情况另行计算，人口发展方程数据流程图如图 1 所示：

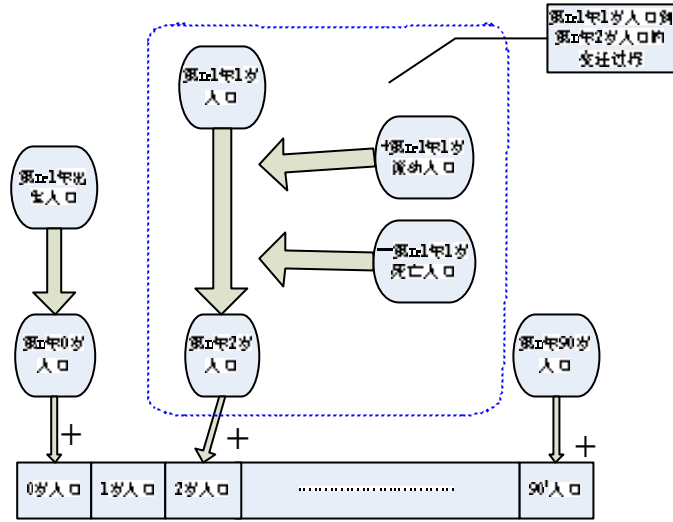


图 1 人口发展方程数据流程图

于是有人口发展方程的离散形式：

$$\begin{aligned}
 f(t) &= b(t) \sum_{i=a1}^{a2} k_i(t) h_i(t) x_i(t); \\
 x_0(t) &= (1 - m_{00}(t)) f(t); \\
 x_1(t+1) &= (1 - m_0(t)) x_0(t) + g_0(t); \\
 x_2(t+1) &= (1 - m_1(t)) x_1(t) + g_1(t); \\
 &\dots\dots \\
 x_m(t+1) &= (1 - m_{m-1}(t)) x_{m-1}(t) + g_{m-1}(t);
 \end{aligned}$$

其中, $k_i(t)$ 是女性人口在该年龄组中的比例, $h_i(t)$ 为生育模式, $b(t)$ 为总和生育率。 $b(t) k_i(t) h_i(t) x_i(t)$ 是 t 年代年龄为 i 的所有妇女在一年中所生婴儿数。 $[a1, a2]$ 为育龄期, $a1 = 15, a2 = 49$ 。这是一个以年度为时间间隔的差分方程组, 并且考虑了婴儿死亡率 $m_{00}(t)$ 的影响。令 $b_i(t) = (1 - m_{00}(t))(1 - m_0(t)) k_i(t) h_i(t), i = a1, a1+1, \dots, a2$; 引进向量和矩阵符号：

$$\begin{aligned}
 \mathbf{x}(t) &= \begin{bmatrix} x_1(t) \\ \vdots \\ x_m(t) \end{bmatrix}, & \mathbf{g}(t) &= \begin{bmatrix} g_0(t) \\ \vdots \\ g_{m-1}(t) \end{bmatrix}, \\
 \mathbf{H}(t) &= \begin{bmatrix} 0 & 0 & \dots & 0 \\ 1 - m_1(t) & 0 & & \vdots \\ & 1 - m_2(t) & & \\ \vdots & & \ddots & \\ 0 \dots & & & 1 - m_{m-1}(t) & 0 \end{bmatrix} \\
 \mathbf{B}(t) &= \begin{bmatrix} 0 \dots 0 & b_{a1}(t) \dots b_{a1}(t) & 0 \dots 0 \\ & 0 & \end{bmatrix}
 \end{aligned}$$

则差分方程组可以写成向量形式：

$$\mathbf{x}(t+1) = \mathbf{H}(t)\mathbf{x}(t) + \mathbf{b}(t)\mathbf{B}(t)\mathbf{x}(t) + \mathbf{g}(t)$$

式中矩阵 $\mathbf{H}(t)$ 为从 t 到 $t+1$ 年的人口状态转移矩阵， $\mathbf{B}(t)$ 为生育矩阵， $\mathbf{g}(t)$ 为人口迁移数。

5.3.3. 参数预测

1. 状态转移矩阵 $\mathbf{H}(t)$

由模型要求，可对 $\mathbf{H}(t)$ 做如下估计：

- i. 由于表中所给的死亡率 $h_i(t)$ 是后向死亡率，需要将其转换成前向死亡率 $m_i(t)$ 才能用于人口发展方程模型，转换方程为： $1 - m_i(t) = 1 / (1 + h_i(t))$ ；
- ii. 考虑数据误差和生物生存规律，对人口按 0, 1~4, 5~9, 10~14, ……85~89, 90 分为 20 组；分别计算组内死亡率，同组内死亡率取相同值，以得到较为稳定的分布规律，
- iii. 其中，年龄 i 的组内死亡率 $h_i(t) = \frac{\sum(\text{男性比例} \times \text{死亡率} + \text{女性比例} \times \text{死亡率})}{\sum(\text{男性比例} + \text{女性比例})}$
- iv. 以上计算得到从 t 到 $t+1$ 年的人口状态转移阵 $\mathbf{H}(t)$ ，考虑到随着生活水平的提高，人们的死亡率会逐渐降低，而儿童和老年人的死亡率受此影响较大，中年人较小，故可由下式预测各年龄段的死亡率

$$m_i(t+1) = \begin{cases} (1 - 1/100)m_i(t) & 0 \leq i \leq 14 \\ (1 - 1/1000)m_i(t) & 15 \leq i \leq 59 \\ (1 - 1/100)m_i(t) & 60 \leq i \end{cases}$$

表 4 各年龄组的死亡率(1: 1000)

| 年龄 | 0 | 1~4 | 5~9 | 10~14 | 15~19 | 20~24 | 25~30 | 31~34 | 35~39 | 40~44 |
|------|-------|-------|-------|-------|-------|-------|-------|-------|--------|--------|
| 2001 | 5.58 | 0.70 | 0.64 | 0.41 | 0.41 | 0.66 | 0.45 | 0.91 | 1.13 | 1.79 |
| 2002 | 11.59 | 0.24 | 0.49 | 0.31 | 0.36 | 0.26 | 0.58 | 0.71 | 1.01 | 1.65 |
| 2003 | 5.17 | 0.19 | 0.21 | 0.21 | 0.12 | 0.21 | 0.45 | 0.89 | 1.24 | 1.75 |
| 2004 | 2.56 | 0.40 | 0.15 | 0.31 | 0.20 | 0.59 | 0.55 | 0.67 | 0.82 | 1.31 |
| 2005 | 6.24 | 0.47 | 0.27 | 0.22 | 0.33 | 0.37 | 0.43 | 0.62 | 0.93 | 1.45 |
| 年龄 | 45~49 | 50~54 | 55~59 | 60~64 | 65~69 | 70~74 | 75~79 | 80~84 | 85~90 | 90~ |
| 2001 | 2.08 | 4.84 | 4.79 | 10.93 | 16.94 | 30.52 | 46.67 | 80.43 | 123.71 | 269.14 |
| 2002 | 2.37 | 2.84 | 5.89 | 9.34 | 16.34 | 28.20 | 49.01 | 73.24 | 104.56 | 155.55 |
| 2003 | 2.48 | 2.64 | 6.25 | 8.84 | 16.69 | 28.40 | 43.82 | 82.37 | 131.69 | 252.08 |
| 2004 | 2.36 | 4.24 | 5.95 | 9.06 | 14.77 | 23.96 | 42.49 | 65.48 | 120.17 | 219.25 |
| 2005 | 2.49 | 3.50 | 5.24 | 8.92 | 15.18 | 24.60 | 41.71 | 69.24 | 116.42 | 275.67 |

2. 生育矩阵 $\mathbf{B}(t)$

生育阵 $\mathbf{B}(t)$ 受婴儿出生死亡率阵 $m_{00}(t)$, 0 岁人口死亡率阵 $m_0(t)$, 生育模式阵 $\mathbf{h}(t)$ 和女性在该年龄组中的比例阵 $\mathbf{k}(t)$ 四个方面的影响, 可以表示为: $\mathbf{B}(t) = (1 - m_{00}(t))(1 - m_0(t))\mathbf{k}(t)\mathbf{h}(t)$ 。 $m_{00}(t)$ $m_0(t)$ 数据已知, 下面分别建立对 $\mathbf{h}(t)$ 和 $\mathbf{k}(t)$ 的预测模型。

i. 生育模式阵 $\mathbf{h}(t)$

生育模式, 实际是指育龄妇女在总体平均意义下单位时间内 (通常指一年), 每个育龄妇女平均活产婴儿数具体在每个年龄上生育的分布形式。在社会稳定的情况下, 可以认为它大体服从一个稳定的分布。本文用标准化后的均值来估计生育模式。

以某一地区为例, 首先将各年的妇女生育率累加, 得到总和生育率, 再取各年龄生育率与总和生育率的比值, 得到标准化后该年的生育模式, 最后对 5 年生育模式取平均, 得到生育模式的估计值。其中 2003 年的数据由于统计错误, 暂不考虑。

$$b(t) = \sum_{i=1}^{a-2} b_i(t), \quad h_i(t) = \frac{1}{4} \sum_{t=2001, t \neq 2003}^{2005} \frac{b_i(t)}{b(t)}$$

利用表中数据可得生育模式为:

表 5 生育模式

| | | | | | | | | | | | | |
|------|--------|-------|-------|-------|-------|-------|-------|-------|-------|--------|--------|--------|
| 年龄 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 |
| 生育模式 | 0.01 | 0.32 | 0.40 | 1.67 | 5.29 | 18.99 | 40.41 | 59.37 | 86.82 | 107.10 | 113.35 | 115.39 |
| 年龄 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 |
| 生育模式 | 101.11 | 83.15 | 69.17 | 51.96 | 35.69 | 29.84 | 18.58 | 18.28 | 11.82 | 8.73 | 6.48 | 4.86 |
| 年龄 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | |
| 生育模式 | 3.24 | 2.27 | 1.40 | 0.87 | 0.54 | 0.67 | 0.47 | 0.33 | 0.17 | 0.27 | 0.45 | |

ii. 女性比例矩阵 $\mathbf{k}(t)$

$$\mathbf{k}(t) = \begin{bmatrix} k_1(t) & & & 0 \\ & k_2(t) & & \\ & & \ddots & \\ 0 & & & k_n(t) \end{bmatrix}$$

$\mathbf{k}(t)$ 是一个对角阵, 下一年女性比例矩阵可以近似为:

$$\mathbf{k}(t+1) = \begin{bmatrix} k_0(t) & & & 0 \\ & k_1(t) & & \\ & & \ddots & \\ 0 & & & k_{n-1}(t) \end{bmatrix}$$

其中 $k_0(t)$ 为上一年出生的女婴所占的比例,根据近年统计资料可知出生女性比例一直在较低的水平进行波动,又因为出生女性比例为随机值,因此可以利用计算机进行为持在低水平的随机模拟方法给出 $k_0(t)$ 。

例如,由表中数据,可取市的模拟区间为 $[0.4604,0.4638]$,而对于乡,则是一个较低的水平 $[0.45,0.4527]$,以后各年初生儿的女性比例都可以在相应地区的模拟区间上随机取值。这样既保证了女性比例的波动性,又保证了它的地区差异性。

3. 总和生育率 $b(t)$

总和生育率是一定期限(如某一年)各年龄组妇女生育率的合计数,说明每名妇女按照某一年的各年龄组生育率度过育龄期,平均可能生育的子女数,是衡量生育水平最常用的指标之一。

受国家政策的影响,在不同的时期, $b(t)$ 有不同的取值。由题中所给数据可知,不同地区的总和生育率有较大差距,近些年来,国家生育政策有所放宽。由此,可以预测未来的一段时间内,各地区总和生育率将趋于平衡,在更长的一段时间内,我国将达到总和生育率为 2.1 的人口稳定状态。又根据国家有关部门预测,全国总和生育率在未来 30 年应保持在 1.8 左右,过高或过低都不利于人口与经济社会的协调发展。故可取 $t=2030$ 年时市镇乡的总和生育率都达到 1.8,当 $t=2050$ 年时达到稳定的总和生育率 2.1,取各时间段间 $b(t)$ 为线性变化。以市为例,可以给出如下的总和生育率 $b(t)$ 的分段函数:

$$\begin{cases} \frac{1.8-b(t)}{2030-t} = \frac{b(t)-1.0}{t-2001} & 2001 \leq t \leq 2030; \\ \frac{2.1-b(t)}{2050-t} = \frac{b(t)-1.8}{t-2030} & 2030 \leq t \leq 2050; \end{cases}$$

4. 人口迁移数 $g(t)$

考虑多个地区时,对 $g(t)$ 的预测采用按年龄有比例迁移的方法。对某一年龄 i , 求出第 $(t+1)$ 年相对于第 t 年的迁移比例 $(x_i(t+1)-x_i(t))/x_i(t)$, 并将其标准化,正数表示迁入,负数表示迁出。分别取 2001、2002 与 2004、2005 两组数据的均值,可得迁移率 $s_i(t)$ 。

在社会稳定的状态下,可认为 $s_i(t)$ 不会随时间有太大变化。根据现有资料,由《国家人口发展战略研究报告》,可设各地区每年有 1% 的人口按此比例迁移,则 $g(t)$ 的表达式为:

$$g(t) = 1\% \times s(t)x(t)$$

5. 人口出生高峰

我国人口在 1984~1990 年出现了的三次人口出生高峰,这将导致 2005 年后的十几年中,育龄妇女占总人口的比例增加,进而引起出生人口和总人口的增加。由于出生高

峰与大约 20 年后的育龄妇女比例有直接关系，为了解出生高峰对总人口的影响，我们采用对大约 20 年后育龄妇女比例进行加权预测。由《国家人口发展战略研究报告》中提供的出生高峰图，可以估计所加权值 p 在 1.2 左右。

5.3.4. 模型求解

将上面所得各参数代入人口发展方程，用 MATLAB 求解^[4]可得出以下数据：

1. 市镇乡及全国人口总数

表 6 市镇乡及全国人口预测（万人）

| 年代 | 2005 | 2010 | 2015 | 2020 | 2025 | 2030 | 2035 | 2040 | 2045 | 2050 |
|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 市 | 31337 | 32112 | 32708 | 32936 | 32635 | 32210 | 31877 | 31415 | 30575 | 29481 |
| 镇 | 16907 | 17531 | 18195 | 18634 | 18698 | 18657 | 18672 | 18685 | 18512 | 18191 |
| 乡 | 82471 | 86638 | 91200 | 94219 | 94864 | 95057 | 95641 | 96425 | 96360 | 95517 |
| 总人口 | 130715 | 136282 | 142103 | 145789 | 146197 | 145924 | 146191 | 146525 | 145447 | 143189 |

2. 各年龄段人数

表 7 不同年龄段人口数预测

| 年代 | 2005 | 2010 | 2015 | 2020 | 2025 | 2030 | 2035 | 2040 | 2045 | 2050 |
|------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 儿童（0～14 岁） | 22658 | 23512 | 27375 | 28557 | 24982 | 20877 | 19953 | 22024 | 23297 | 22265 |
| 中年（15～59） | 91480 | 92979 | 90423 | 89911 | 87767 | 85885 | 83517 | 81290 | 78055 | 73886 |
| 老年（60～） | 16576 | 19790 | 24305 | 27321 | 33449 | 39161 | 42721 | 43210 | 44095 | 47039 |

3. 女性在总人口中的比例

表 8 女性人口百分比预测

| 年代 | 2005 | 2010 | 2015 | 2020 | 2025 | 2030 | 2035 | 2040 | 2045 | 2050 |
|------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 女性人数 | 63466 | 65774 | 68205 | 69673 | 69637 | 69267 | 69132 | 69000 | 68205 | 66885 |
| 总人数 | 130715 | 136282 | 142103 | 145789 | 146197 | 145924 | 146191 | 146525 | 145447 | 143189 |
| 女性比例 | 48.55 | 48.26 | 48.00 | 47.79 | 47.63 | 47.47 | 47.29 | 47.09 | 46.89 | 46.71 |

5.3.5. 结果分析与检验

1. 由表 6 可得，全国人口在 2010 年、2020 年的人口将达 13.63、14.57 亿人，同《国家人口发展战略研究报告》中的预测值高度一致，且由图 2 看出，全国人口将呈现在短期内增长、中期平稳过渡、长期缓慢下降的趋势，在 2039 年达到峰值 14.65 亿人，这符合我国的人口政策。

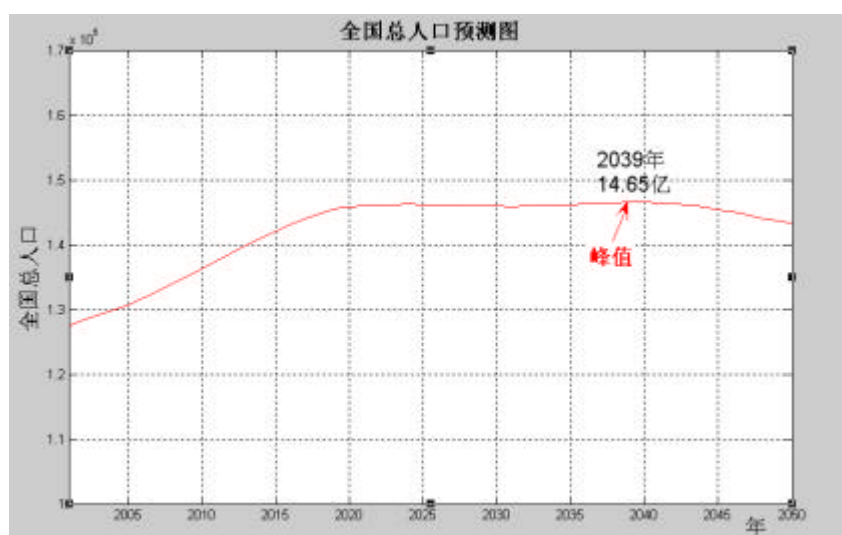


图 2 全国人口预测曲线图

2. 由表 7 看出，随着时间的推移，老年人所占人口比例越来越大，并在 2031 年突破 4 亿大关，占总人口的 27.48%，老龄化问题应当引起足够重视！

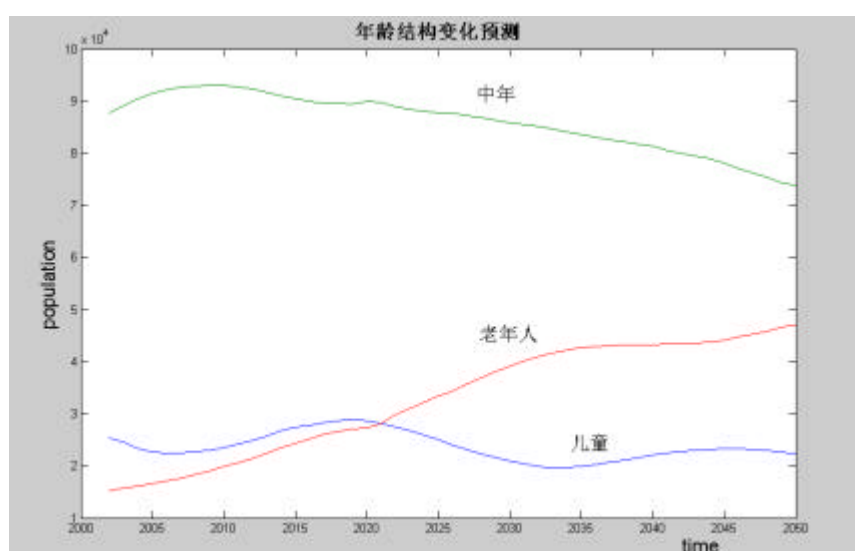


图 3 年龄结构变化预测图

3. 由表 8 看出，我国人口的另一个特点是男女比例失调，若按目前出生人口的性别比持续下去，到 2015 年后，性别比将低于正常水平 48%，新进入婚育年龄人口男性明显多于女性，婚姻挤压问题凸现，可能导致的社会秩序混乱将成为影响社会稳定的严重隐患。
4. 由图 2 可以看出，1984~1990 年间的人口出生高峰期，直接导致了 2005~2015 年的婴儿出生高峰，并间接引起了 2030~2045 间的人口总数波动，且在 2039 年出现了人口总数峰值，说明人口出生高峰将产生不止一代的影响，应尽量避免高峰期的产生。

5.3.6. 敏感性分析

1. 总和生育率 b

由模型的构建过程可知 ,总和生育率对预测结果有着重大影响 ,此点可由图 4 说明 :

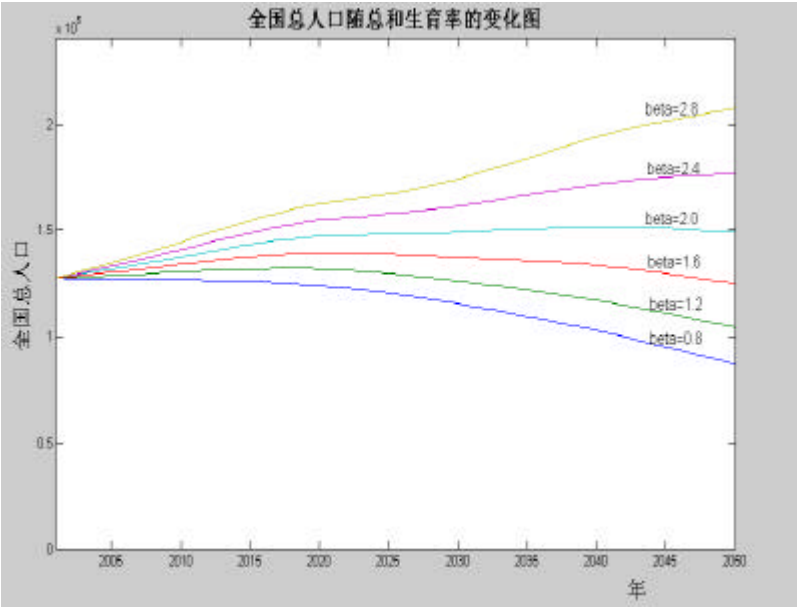


图 4 全国总人口随总和生育率的变化图

由图 4 可知 ,随着 b 的增大 ,总人口数和其变化趋势都将产生较大变化 ,若要维持人口的长期稳定 , b 应在 2.0 左右 ,与 2.1 的更替水平相接近 ;若要控制现阶段人口的过快增长 ,则 b 值应适当低于 2.0。 $b = 1.8$ 的效果可参照图 2 ;可知模型对 b 非常敏感。

2. 迁移率

随着计划生育国策的进一步推广 ,城镇和乡村的总和生育率将趋于一致 ,在此前提下 ,分别取迁移率为 1/100 , 1/80 , 1/50 ,可以得到如下结果 :

表 9 迁移率对总人口的影响

| 迁移率 | 2010 | 2015 | 2020 | 2025 | 2030 | 2035 | 2040 | 2045 | 2050 |
|-------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 1/100 | 136282 | 142103 | 145789 | 146197 | 145924 | 146191 | 146525 | 145447 | 143189 |
| 1/80 | 135950 | 141530 | 144998 | 145225 | 144759 | 144795 | 144880 | 143583 | 141150 |
| 1/50 | 135895 | 141407 | 144818 | 145043 | 144594 | 144651 | 144741 | 143435 | 141000 |

表中数据说明随着迁移率的增加 ,总人口数将降低 ,这同城市的低生育率、低死亡率相符合 ,同时表中总人口数目变化不大 ,也说明模型对迁移率的敏感性不高。

3. 老龄化和出生高峰

老龄化和出生高峰都将对年龄结构产生影响 ,它们主要通过影响某一时段育龄妇女占总人口的比例来引起人口的总数的变化 ,因而我们可以通过研究育龄妇女的比例来对

老龄化和出生高峰的影响进行预测，表 10 是预测结果：

表 10 不同权值下育龄妇女比例对总人口的影响

| 加权因子 | 2010 | 2015 | 2020 | 2025 | 2030 | 2035 | 2040 | 2045 | 2050 |
|-------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| p=1.1 | 135145 | 139845 | 142679 | 142922 | 142332 | 141927 | 141392 | 139639 | 137026 |
| p=1.2 | 136282 | 142103 | 145789 | 146197 | 145924 | 146191 | 146525 | 145447 | 143189 |
| p=1.3 | 136820 | 143334 | 147478 | 147720 | 147410 | 147933 | 148685 | 147879 | 145647 |

由表 10 中的数据可知，随着育龄妇女加权因子的减小，也及老年化突出程度增加和出生高峰严重程度的减少，总人口数将减少，说明模型对老龄化和出生高峰的处理符合实际情况。同时可以看出，它们对总人口的影响是比较大的，因而模型对老龄化和出生高峰比较敏感。

4. 性别比例

联合观察表 6 和表 8 中的数据可以对性别比例对总人口的影响做大致了解，从表 11 可以看出，总人口随着女性比例的降低而降低，符合生物学规律。

表 11 总人口与女性比例（%）关系表

| | 2010 | 2015 | 2020 | 2025 | 2030 | 2035 | 2040 | 2045 | 2050 |
|------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 女性比例 | 48.26 | 48.00 | 47.79 | 47.63 | 47.47 | 47.29 | 47.09 | 46.89 | 46.71 |
| 总人口 | 136282 | 142103 | 145789 | 146197 | 145924 | 146191 | 146525 | 145447 | 143189 |

5.3.7. 模型优缺点

模型优点：

1. 充分考虑现阶段的人口增长特点，引入性别比例转移矩阵，利用计算机模拟未来出生人口的性别比例，动态地对人口性别变化做出预测。
2. 对死亡率做出合理的假设，分年龄段地计算各年龄的死亡率，既简化了模型，也较准确地符合实际。
3. 结合我国当今大量人口迁入城市以及城镇化进程快速，将人口迁移率引入模型，得到各年市、镇、乡的各年人口数。
4. 八十年代的出生高峰期会使最近几年的总和生育率提高，模型通过引入权重，使模型更贴近实际。

模型缺点：

1. 总和生育率是一个随时间变化的量，本文没有对其变化的因素充分考虑，只是按照国家统计局的数据和国家的政策对其变化规律做了线性近似。
2. 生育模式在社会稳定时，一般不会变化，但对于长期的人口预测，生育模式应该变化。
3. 死亡率是一个随机变量，本模型没有对其随机性进行展开讨论。
4. 没有考虑女性平均寿命比男性高对老年人中女性比例的影响。

6. 两个模型的比较

6.1. 复杂程度

模型一基于传统的 logistic 模型，所需数据少，具有简洁易行等特点，且由于用差分方程代替了微分方程，使求解难度大大降低。而模型二构造复杂，所需数据种类多，量大，求解也较复杂，需要用到专门的数学软件。

6.2. 预测精度

由模型得到的数据可以看出，对于中短期的预测，两模型均有较高的准确度，但随着时间的增加，模型一的精确度逐步下降，而模型二则很好的维持了高精度。

6.3. 应用范围

由以上分析可知，两个模型各具优点，若拥有的数据量有限，且只需进行短期预测，首选模型一；若拥有足够的数据量，且需对人口进行长期预测，则可以考虑模型二。

7. 对政府的一些建议

人口数量和素质与国家的发展息息相关，根据合理的预测方案采取正确的人口措施，对于国家政府至关重要，这里我们基于以上两个模型提出一些建议：

1. 从模型的灵敏度出发，总和生育率对人口的增长有很大影响，因此应严格控制总和生育率，有关部门要及时搜集精确数据，对总和生育率做出准确判断，采取相应措施。
2. 由模型知，男女比例对人口总数也有一定的影响，且有可能导致社会秩序混乱是影响社会稳定的严重隐患，应引起足够重视。
3. 出生高峰将导致某段时间内人口总数的快速增长，对本已稳定的社会关系造成冲击，且将对后几代产生持续影响，因此在人口政策中应避免出现出生高峰。

8. 参考文献

- [1] 国家统计局编委会，中国人口统计年鉴 2006，北京：中国统计出版社，2006。
- [2] 姜启源 谢金星，数学模型，北京：高等教育出版社，2003。
- [3] 宋健 于景元，人口控制论，北京：科学出版社，1985。
- [4] 苏金明 张莲花 刘波等，MATLAB 工具箱应用，北京：电子工业出版社，2004。