

# 中国人口发展的预测模型

## 摘要

本文就中国人口的总数及组成问题进行了研究并分别建立了线性模型，Logistic 模型、多元线性回归模型，考虑年龄结构而构建的 Leslie 矩阵预测模型四个模型。

Logistic 模型和线性模型根据近十年中国人口总数的统计数据，用最小二乘法确定参数，并进行统计检验，证明可较为精确地预测短期未来几年的中国人口总数变化。

接着我们以 Leslie 矩阵构建了人口的动力学方程，建立了 2001 年以来中国人口发展的数学模型。本文一大特色是：所用的 Leslie 矩阵更为科学——年龄别生育率是随时间变化的，并由此得到时变的 Leslie 矩阵。（注：题目所给的 2003 年的年龄别出生率明显异常，我们通过插值处理后将其求出。而年龄别死亡率随时间的变化本文认为可忽略，正文中有图解说明。）利用该模型可推算较长一段时间内的随年以岁为单位的年龄构成。

多元线性回归模型来自多元统计中的相关知识，该模型的特点是综合考虑了出生率、死亡率和性别比三个因素对人口总数的影响。我们还论述了性别比失调将对我国，尤其是农村人口的影响，并提出了几条改进建议。整个计算过程采用 SPSS 软件，一次性输出多元线性回归的全部计算结果。

我们对四个模型进行了比较和评价，并用人口普查的数据验证了模型的有效性。最后进行了优缺点分析。

在“模型扩展”部分，我们又对老龄化与人口红利问题，“二胎政策”等我国人口形势面临的紧迫问题进行了研究和预测。本文所建模型有很好的移植性，理论上还可通过调整参量定量分析一部分人口政策及社会因素对人口发展的影响，可供有关研究及政策制定部门参考。

## 关键词

线性模型   Logistic 模型   多元线性回归模型   Leslie 矩阵预测模型   老龄化   人口红利   二胎政策



## 本文目录

摘要.....	1
关键字.....	1
问题重述.....	3
基本假设.....	3
符号说明.....	3
模型的建立与求解.....	4
一. Logistic模型和线性回归模型.....	4
二. Leslie基本模型: .....	5
三. 多元线性回归模型: .....	10
模型扩展.....	11
一、老龄化与人口红利问题.....	11
二. 二胎政策.....	13
模型灵敏度分析.....	15
模型评价.....	16
一. 模型对比与分析: .....	16
二. 模型优缺点: .....	18
参考文献.....	20



## 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。根据已有的数据，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。

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。但值得注意的是，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（以及从人口红利向人口负债的转变问题），出生人口性别比持续升高，还有乡村人口城镇化等因素，这些都很大程度上影响着中国人口的增长。所以，我们此次建模并不是“老生重谈”，而是与时俱进地以新视角对未来我国的人口状况进行全面的分析。

本文试从中国的实际情况和人口增长的上述特点出发，并参考了相关数据，建立了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的趋势做出了预测，本文最后还指出了模型中的优点和不足。

## 基本假设

1. 目前我国的移民现象很少见，我们可以认为中国人口是一个封闭的系统；
2. 假设一个家庭只生一个孩子；
3. 不考虑地区差异；
4. 由于当年龄相差不大时，生育率指标非常接近可视为常量，因此将育龄期女性分成7个年龄组，其中15-19岁为第1个年龄组，20-24岁为第2个年龄组，25-29岁为第3个年龄组，30-34岁为第4个年龄组，……，45-49岁为第7个年龄组；
5. 假设在本文所预测的时期内，不会有突发性的，不可避免的客观因素使人口数量有较大波动；
6. 本文中死亡率和生育率均为十万分比；

## 符号说明

$x_i(k)$	时段 $k$ 第 $i$ 年龄组的人口数量
$x(k)$	$k$ 时段总人口的数量（不考虑年龄）
$b_i$	第 $i$ 年龄组的生育率（该年龄组某年的出生人数与该年龄组该年总人数之比）
$d_i$	第 $i$ 年龄组的死亡率（该年龄组某年的死亡人数与该年龄组该年总人数之比）
$s_i$	第 $i$ 年龄组 1 个时间段的存活率，则 $s_i = 1 - d_i$
$\bar{x}(k) = [x_1(k), x_2(k), \dots, x_n(k)]^T$	时段 $k$ 人口按年龄组的分布向量
$x_i^{rural}(k)$	乡村时段 $k$ 第 $i$ 年龄组的人口数量



$x_i^{urban}(k)$	城镇时段 k 第 i 年龄组的人口数量
$x_i^{city}(k)$	城市时段 k 第 i 年龄组的人口数量
$\eta^{rural}, \eta^{urban}, \eta^{city}$	乡村, 城镇, 城市人口占总人口的百分比
$r$	人口固有增长率
$x_m$	人口容量, 即自然资源和环境条件所能容纳的最大人口数

## 模型的建立与求解

### 一. Logistic 模型和线性回归模型

#### 1. Logistic 模型:

通常由于自然资源, 环境条件等因素的影响, 人口数不能持续无限的增长, 长到一定数量后, 增长率会下降, 并且随着人口的增加, 环境的组织作用会越来越大。在该模型中, 我们忽略年龄对增长率的影响。

该阻滞作用体现在对人口增长率  $r$  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  $r(x)$ , 则它应该是减函数。于是得

$$\frac{d_x}{d_t} = r(x)x \quad (1)$$

$$x(0) = x_0 \quad (2)$$

对  $r(x)$  的一个最简单的假定是, 设  $r(x)$  为  $x$  的线形函数, 即

$$r(x) = r - sx \quad (r < 0, s > 0) \quad (3)$$

当  $x = x_m$  时, 人口不再增长, 即增长率  $r(x_m) = 0$ , 代入③式得  $s = \frac{r}{x_m}$ , 于是③

式为

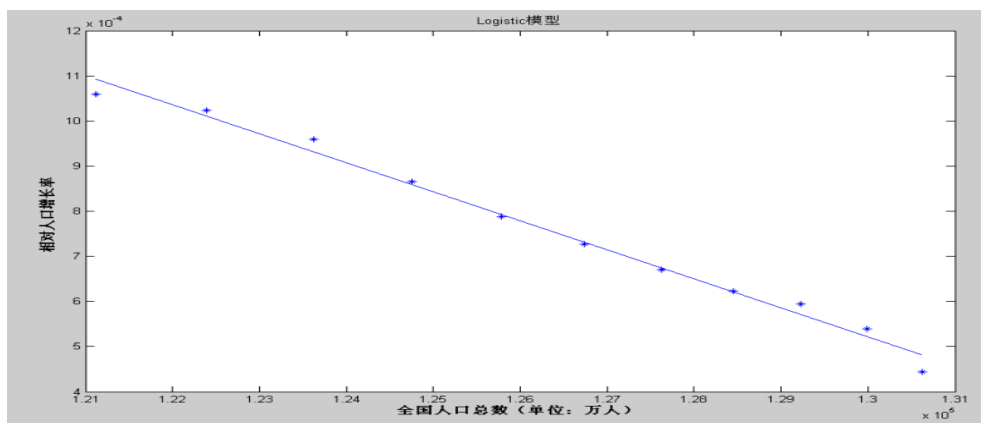
$$r(x) = r(1 - \frac{x}{x_m}) \quad (4)$$

④式的一种解释为, 增长率与人口尚未实现部分的比例成正比, 比例系数为固有增长率  $r$ 。

用 MATLAB 软件计算得到  $r=0.8883\%$ ,  $x_m=13.8114$  亿。

下图为  $r(x)$  关于  $x$  的线形函数图



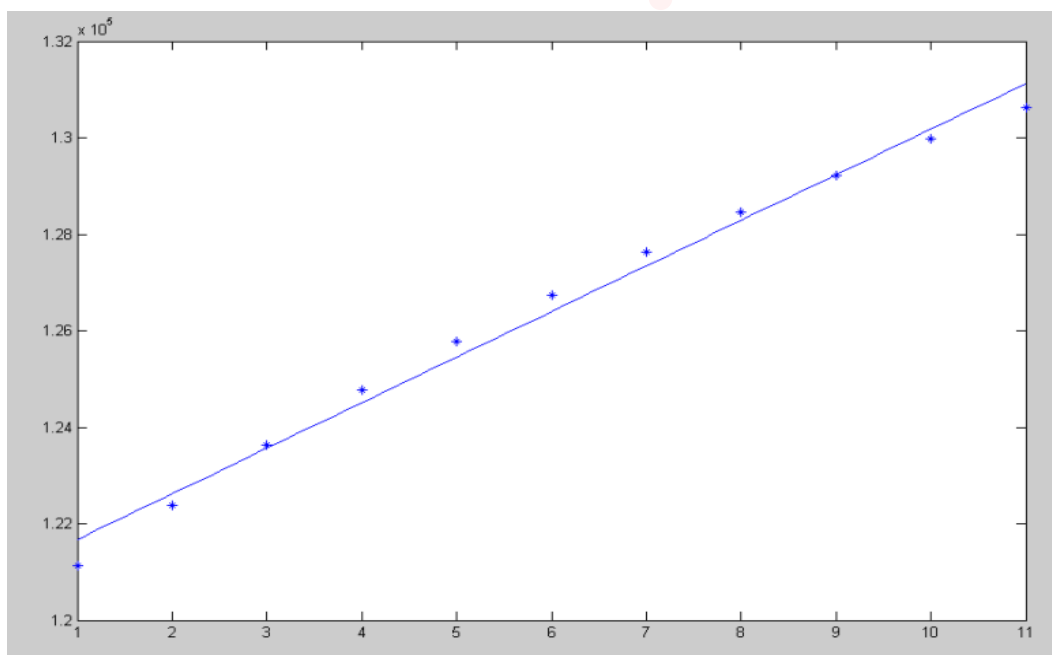


## 2. 线性模型:

在该模型中, 我们假设人口数量随时间线性增加, 即

$$x(k) = ak + b \quad (5)$$

我们利用 1995—2005 年的人口数据, 用最小二乘法, 可以计算出参数  $a=945.081818182$   $b=120724.872727273$ 。如图:



决定系数 $R^2$	F 值	P 值	a 的置信区间	b 的置信区间
0.9895	846.9792	<0.0001	[870, 1020]	[120230, 121220]

决定系数 0.9895 说明 a, b 的决定作用很大, 拟合结果很好。

## 二. Leslie 基本模型:

### (一) Leslie 矩阵

生育率  $b_i$  和存活率  $s_i$  可以通过处理所给的数据计算得到。每年龄段  $x_i(k)$  的变化规律



由以下的基本事实得到：时段  $k+1$  第 1 年龄组的人口数量是时段  $k$  各年龄组生育的人口数量之和，即

$$x_1(k+1) = \sum_{i=1}^n b_i x_i(k) \quad (1)$$

时段  $k+1$  第  $i+1$  年龄组的人口数量是时段  $k$  第  $i$  年龄组存活下来的数量，即

$$x_{i+1}(k+1) = s_i x_i(k) \quad , \quad i=1, 2, 3, \dots, n-1 \quad (2)$$

记时段  $k$  人口按年龄组的分布向量为

$$\bar{x}(k) = [x_1(k), x_2(k), \dots, x_n(k)]^T \quad (3)$$

由生育率  $b_i$  和存活率  $s_i$  构成的 Leslie 矩阵 (以下简称  $L$  矩阵) 如下：

$$\bar{L}_k = \begin{bmatrix} b_1 & b_2 & \dots & b_{n-1} & b_n \\ s_1 & 0 & & 0 & 0 \\ & & \ddots & & \\ & s_2 & & 0 & \vdots \\ & & \ddots & & \\ 0 & & & s_{n-1} & 0 \end{bmatrix} \quad (4)$$

则 (1), (2) 可以表示为

$$\bar{x}(k+1) = \bar{L}_k \bar{x}(k) \quad k=0, 1, 2, \dots \quad (5)$$

当矩阵  $\bar{L}$  和按年龄组初始分布向量  $\bar{x}(0)$  已知时，可以预测人以时间段  $k$  的种群按年龄组的分布为

$$\bar{x}(k) = \prod_{i=0}^k \bar{L}_i \bar{x}(0) \quad k=1, 2, \dots \quad (6)$$

有了  $\bar{x}(k)$  当然不难算出时段  $k$  人口的总数。

## (二) 参量的确定

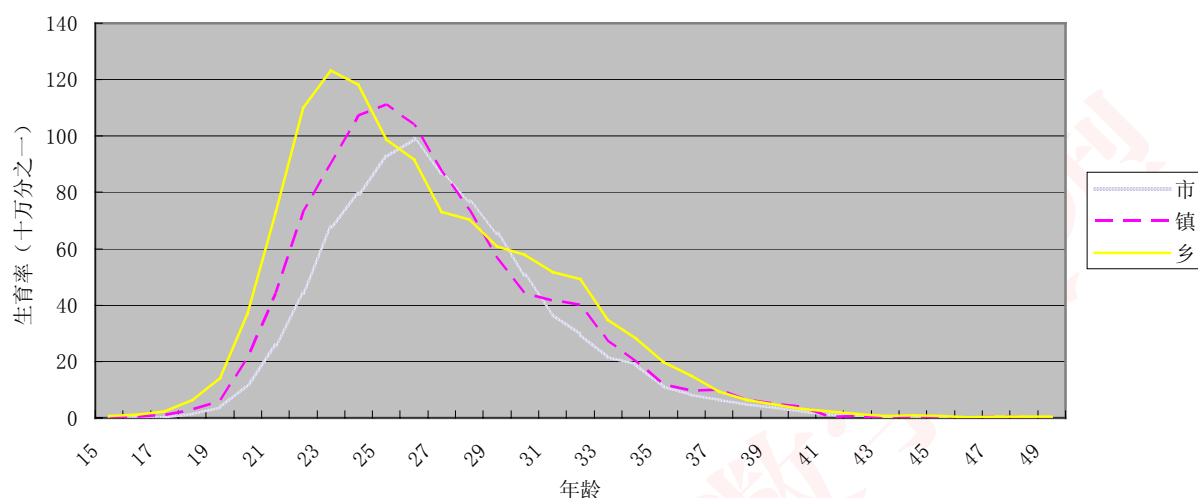
LESLIE 矩阵本身是普适的人口动力学方程，而不同历史，社会条件下的人口发展模式特点是由其中的参量来描述的，也就是  $b_n(i)$  和  $d_n(i)$ 。它们都是随时间变化的量，是诸多因素共同作用的结果。我们不可能找到这些参量每年的精确数值，因此作一些假设和近似是必要的。

### 1. 年龄别生育率

我们首先考察城市，镇和乡村人口的生育的年龄结构，我们发现三者有很大的不同。我们把某个时点的妇女的生育年龄分布同时看作一个妇女一生中生育孩子数的分布。



市，镇，乡女性生育率年龄分布



从图中，我们发现乡村妇女的总和生育率明显高于城市妇女，同时她们的生育高峰也比城市妇女早一些，且计划生育政策及其执行情况也有很大的不同，所以本文中把这两种情况分开讨论。将市，镇，乡生育率进行加权，求出全国的生育率。其中权重为市，镇，乡人口占全国人口的比例，即

$$x_{i+1}(0) = \eta^{rural} x_{i+1}^{rural}(0) + \eta^{urban} x_{i+1}^{urban}(0) + \eta^{city} x_{i+1}^{city} =$$

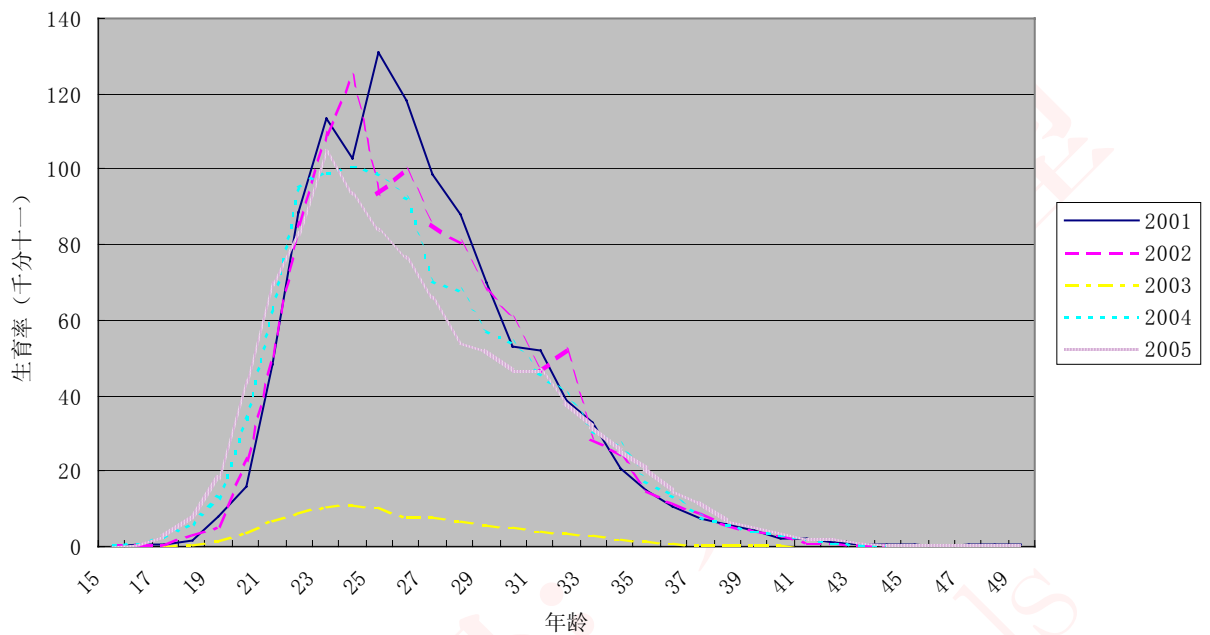
$$\left[ \sum_k \eta^{rural} b_i^{rural}(i) x_i^{rural}(k) + \sum_k \eta^{urban} b_i^{urban}(k) x_i^{urban}(k) \right. \\ \left. + \sum_k \eta^{city} b_i^{city}(k) x_i^{city}(k) \right] / 2$$

知道了  $x_i(k)$ ，我们可以用城乡人口比来得到  $A_i^{rural}(k)$  和  $A_i^{urban}(k)$ ， $A_i^{city}(k)$ 。

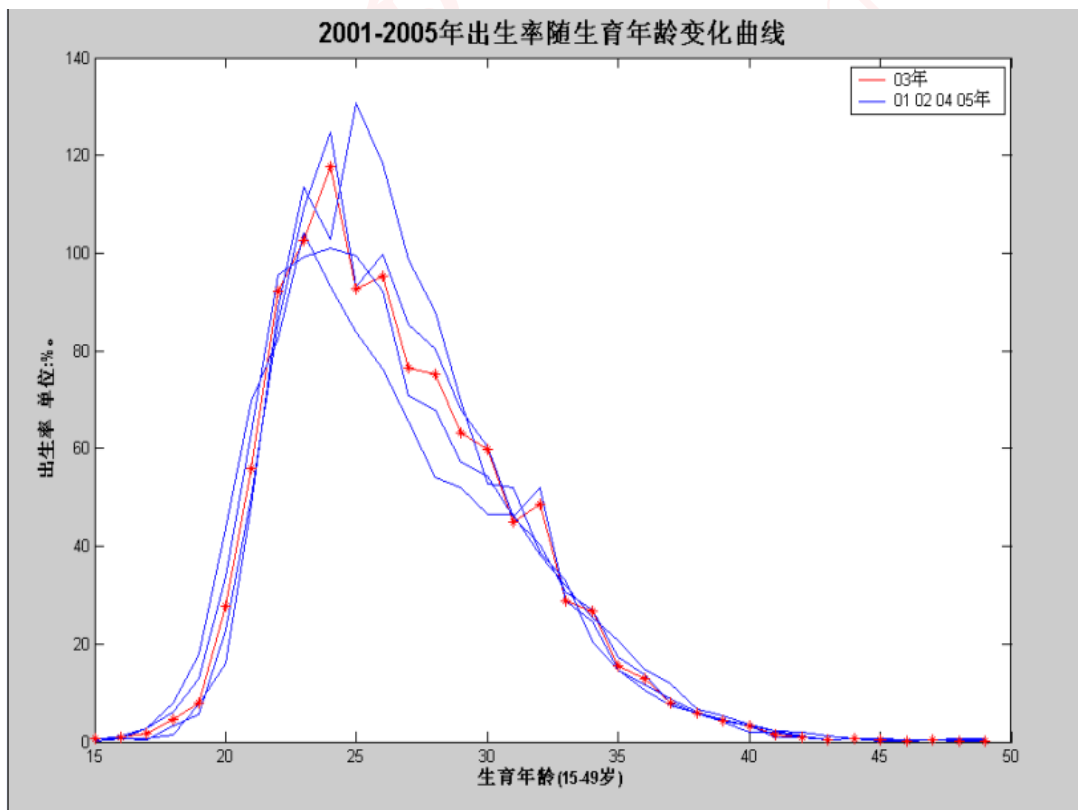
然后，我们将中国人口按年龄大小，等间隔地分成  $n$  个年龄组。我们就以 1 岁为一个年龄组，以年为一个时间段。我们将 2001~2005 年的生育率随年龄变化曲线绘制在同一张表格中，发现 2003 的曲线明显异常。如图：



全国生育率

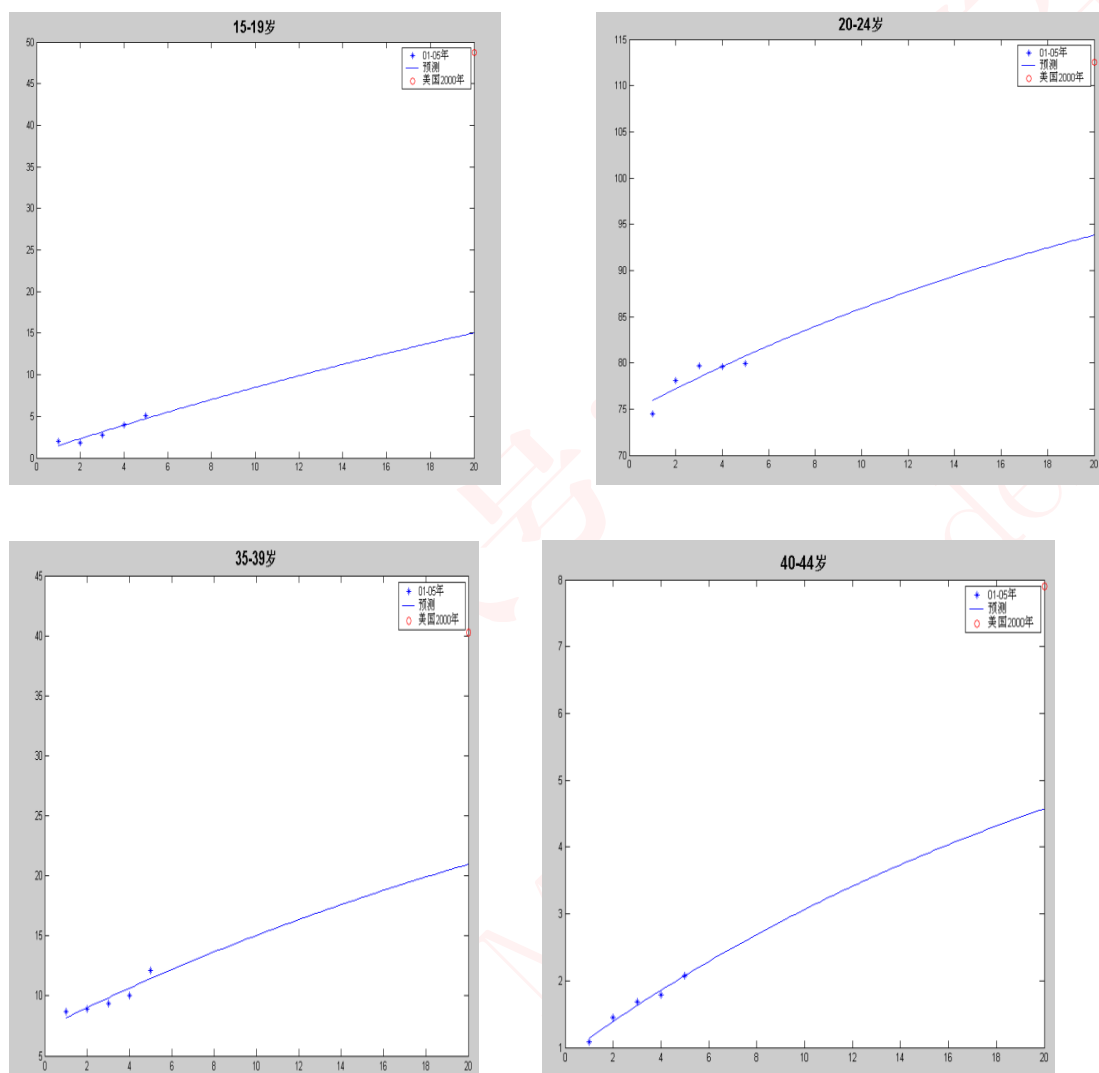


于是我们根据其他 4 年的曲线插值出了 2003 年的曲线，如下图所示，我们认为插值得到的曲线能够较好地反应 2003 年的生育率情况。





由图可以看出，出生率在各年份差别比较大，即出生率随时间  $k$  的变化而变化。由于当年龄相差不大时，生育率指标非常接近可视为常量，因此我们将育龄期女性分成 7 个年龄组，其中 15-19 岁为第 1 个年龄组，20-24 岁为第 2 个年龄组，25-29 岁为第 3 个年龄组，30-34 岁为第 4 个年龄组，……，45-49 岁为第 7 个年龄组。我们用发达国家（美国）2000 年各年龄段女性的生育率作为中国生育率的为定值，且假设经过 20 年可以达到稳定，由指数模型拟合出各年龄段女性的生育率随时间的变化曲线，再由此计算不同年龄组妇女的生育率。

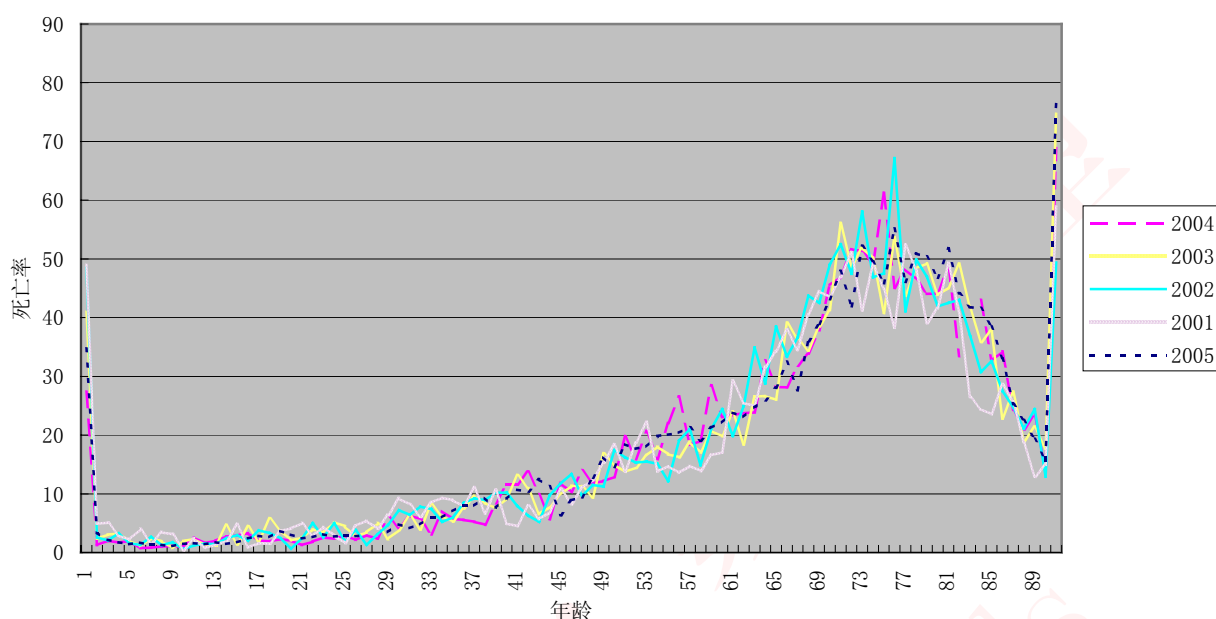


	2001	2002	2003	2004	2005	美国
15-19	2.044953118	1.833445218	2.732876392	3.924706246	5.036612969	48.7
20-24	74.45850832	78.0876079	79.68058037	79.6183235	79.8962862	112.5
35-39	8.658083668	8.894663124	9.381964459	10.04661855	12.09383297	40.3
40-44	1.080822096	1.44612788	1.681532414	1.79416526	2.081224771	0.5

## 2. 年龄别死亡率 $d_i(k)$



死亡率随年龄变化曲线



通过计算机绘图分析，如上图所示，各条曲线比较接近，所以我们可以合理地假设死亡率  $d_i$  不随时间段  $k$  的变化而变化（从而存活率  $s_i$  是不随时间段  $k$  变化的），只与年龄有关。即使在未来的很长一段时间内，只要没有医学上大的革命，这个量的变化将很小。因此，本文中把它当成时间的常量处理。于是我们取各年份的死亡率的平均值作为死亡率  $d_i$ 。（注：我们对原始数据中的 2 处异常点做了插值处理，分别为：2004 年 82 岁年龄段的死亡率和 2005 年 27 岁年龄段的死亡率）

### 三. 多元线性回归模型：

设时间顺序为自变量  $X_1$ ，出生率为自变量  $X_2$ ，死亡率为自变量  $X_3$ ，性别比率为自变量  $X_4$ ，年底人口数为因变量  $Y$ 。把表 1 中的第一列年份与时间顺序对应，调查样本容量 5，即  $N=5$ 。采用多元统计中的多元线性回归方法，针对表 1 中的调查数据，整个计算过程采用 SPSS 软件，一次性输出多元线性回归的全部计算结果。

表 1

年份	时 间 顺 序	出生率	死亡率	性比率（男/女）	年底人口数 （万人）	绝对误差	相对误差
2001	1	13.38	6.43	106.00	127627.0	-125.3559	-0.0010
2002	2	12.86	6.41	106.10	128453.0	58.1461	0.0004
2003	3	12.41	6.40	106.20	129227.0	235.9285	0.0018



2004	4	12.29	6.42	106.30	129988.0	267.9274	0.0021
2005	5	12.40	6.51	106.30	130756.0	-139.6701	-0.0011

假设因变量  $Y$  的观测值具有相同的方差，并服从正态分布，把自变量  $X_1, X_2, X_3, X_4$  和因变量  $Y$  的样本数据输入 SPSS 中，计算得回归平方和 400986.51；方程复相关系数 0.998，决定系数 0.996，方差分析的  $F=374.045$ ， $F$  值对应的最小概率  $P=0.000$ 。若显著水平取 0.01，则多元线性回归方程为：

$$Y = 164551.99 + 942.172X_1 + 119.794X_2 - 6610.520X_3 - 25.875X_4$$

## 模型扩展

### 一、老龄化与人口红利问题

#### 1. 生育政策：

通过对生育率，死亡率和老龄化之间关系的分析，我们发现，生育率下降是引起和加速人口老化的主要因素。而我国生育率水平的下降主要是实行计划生育政策而实现的。

为了阐明生育政策对未来人口的影响和人口压力对生育政策选择的影响，我们将对不同生育政策方案的人口预测结果进行分析。基于我国目前执行的生育政策和实际生育率水平，我国生育政策将有如下几种选择：

政策选择 1：继续执行现行低生育率水平的生育政策，并使总和生育率在 21 世纪初实现政策要求的水平（ $TFR = 1.6$ ）。

政策选择 2：继续执行现行低生育率水平的生育政策，并保持目前的总和生育率水平（ $TFR = 1.8$ ）。

政策选择 3：执行选择一定时机将总和生育率调整到更替水平并将保持不变的生育政策。

政策选择 4：执行较高生育率水平的生育政策（ $TFR > 2.10$ ）。

依据上述几种可选择的生育政策，我们拟定表 1 所示 6 个对应的总和生育率变化模式。

表 1 中国人口发展趋势预测的六种方案的总和生育率（TFR）假设

	1990-1995	1995-2000	2000-2005	2005-2010	2010-2015	2015-2020	2020-2025	2025-2030	2030-2050	2050-2100
1	2.2	1.9	1.8	1.6	1.6	1.6	1.6	1.6	1.6	1.6
2	2.2	1.9	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8
3	2.2	1.9	1.8	1.8	1.8	1.8	1.8	1.8	1.8	2.1
4	2.2	1.9	1.8	1.8	1.8	1.8	1.9	2.0	2.1	2.1
5	2.2	1.9	2.1	2.1	2.1	2.1	2.1	2.1	2.1	2.1
6	2.2	1.9	2.1	2.2	2.3	2.3	2.3	2.3	2.3	2.3

其中，方案 1 是对应于政策选择 1；方案 2 对应于政策选择 2；方案 3、4、5 对应



于政策选择 3 ;方案 6 对应于政策选择 4。

## 2. 人口规模预测结果分析

依据上述方案的设定,对我国人口整个 21 世纪进行超长趋势预测。另外,依据以往研究,我国资源环境的人口承载极限在 16 亿左右。下面我们将 16 亿人口作为我国资源环境的人口承载极限,也就作为我们人口控制的目标,来决定我们对上述几个方案的取舍。

人口预测结果显示,方案 6 的人口处于持续增长的状态,到 2035 年人口超过 16 亿,2100 年接近 20 亿,显然按该方案我国人口将超出承载极限。方案 5 的人口也是处于持续增长的状态,2035 年人口超过 16 亿,但从 2055 年后开始增长缓慢,表现为零人口增长状态,到 2100 年人口总量为 17.2 亿,也超出了承载极限。因此方案 5 和 6 均不可取。

方案 1 和方案 2 的人口分别于 2030 年、2035 年达到峰值,分别为 14.55 亿和 14.99

亿。此后,人口将处于负增长状态,到 2100 年分别减少为 9.36 亿和 11.88 亿。按这两个政策方案和拟定的总和生育率水平虽然可以使人口控制在 16 亿以内,而且逐渐减少,能够减缓人口规模庞大所造成的压力,但从长远的人口发展看,人口将最终自行消亡,因此这两个政策是不可采取的。由此看来,低生育率水平不能成为我们永远的政策选择,我们必须选择适当的时机将总和生育率水平控制在更替水平。那么,选择什么时间和时机将生育率水平调整到更替水平是我们作出的政策选择。

方案 3 的零人口增长点在 2035 年,人口峰值为 14.99 亿,此后在未来 65 年里处于负增长,于 2100 年实现零人口增长域,最终人口规模为 13.8 亿;方案 4 的零人口增长点在 2045 年,人口峰值为 15.34 亿,此后经历 40 年的缓慢负增长,于 2085 年再次经过零人口增长点后人口进入零人口增长域,其间人口规模在 15.1 亿以上。通过这里的描述,方案 3 和 4 都符合人口控制的目标。

如果我们仅依靠人口规模作出判断,无疑政策选择方案 3 是最佳的。但是,政策的选择恰恰不能仅依靠人口规模,还要考虑到人口年龄结构的合理性和劳动力的供给。

表 2 不同生育水平下的老年人口比重 (%)

	2000 年	2030 年	2050 年	2100 年	老年人口比重最大值	
					年份	比重
1	7.1	15.8	23.0	29.1	----	----
2	7.1	15.4	21.6	25.9	----	----
3	7.1	15.4	21.4	22.3	2065	23.5
4	7.1	15.3	20.6	21.5	2065	22.2
5	7.1	14.6	19.0	20.3	----	----
6	7.1	14.6	18.9	19.3	2060	19.6

在表 2 中我们给出各方案的老年人口比重,从预测结果显示,方案 4 人口的老年人口比重在 2045 年以后(除 2060—2065 年的 5 年时间)一直稳定在 22 %以下,是我们无论如何总要面对的比重。然而方案 3 人口在进入零人口增长域以前老年人口比重在一相当长的时期(2055—2100 年)里会达到 22—23.5 %。我们还不能解决得了这样庞大的老年人口所带来的问题。因此,我们认为在人口规模控制在 16 亿以内的情况下,选择老年人口比重较小的方案 4 是明智的。

通过以上分析表明,在 21 世纪考虑到人口老化问题制约,生育政策应在实现零人口增长点以前进行调整(于 2030 年)使总和生育率达到更替水平。

## 3. 老龄化与人口红利问题



根据预测，随着中国总人口在 2033 年达到 14.29 亿之前继续增加，抚养比首先将进一步下降，从 2000 年的 42.6% 下降到 2015 年的 39.4%，下降 3.2 个百分点，带来经济增长率上升 0.4%。这个阶段大约持续到 2015 年前后。此后，随着人口老年化速度上升，人口转变对经济增长的贡献将由“人口红利”阶段转为“人口负债”阶段，人口老龄化加速将给中国经济长期增长带来一定程度的负面影响。因此，2015 年前后是中国“人口红利”阶段的转折点。

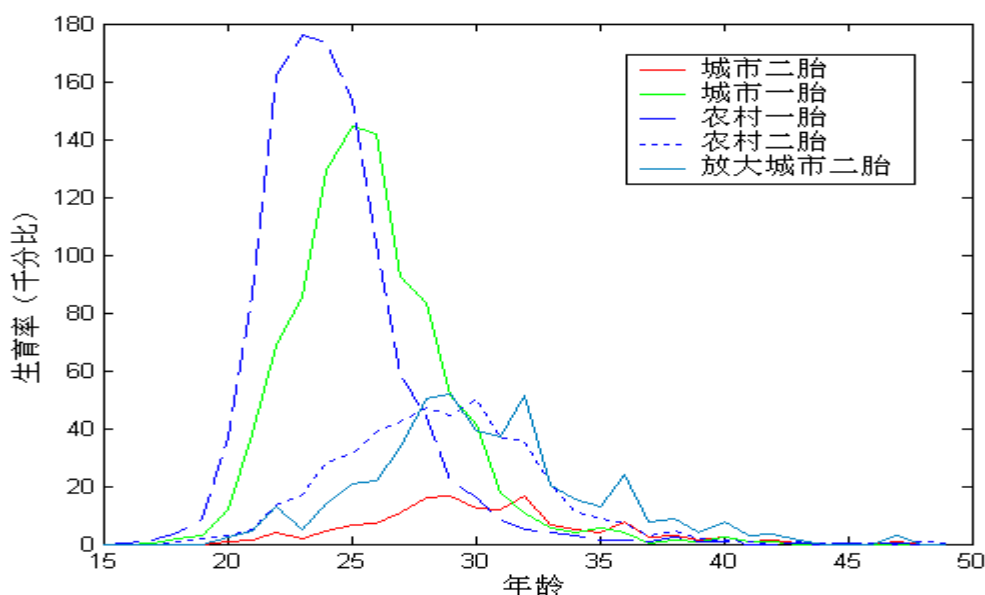
实际上，劳动年龄人口的增长速度已经下降。按照目前的预测，如果没有农村劳动力向城市的转移，最迟在 2009 年，非农产业发展所需的劳动力就要告罄。预测还显示，少儿抚养比下降与老年抚养比上升之间的间隔将很短，使得总体抚养比只在相对短暂的时期内处于较低的水平上，即在本世纪只会再继续一个较短的下降，最低点为 2013 年的 38.8%，随后将以老年抚养比提高为特征大幅度回升；其结果是社会总体养老负担迅速加重。到 2017 年，老年人口占总人口的比例将超过 10%，在少儿抚养比仍然高达 26.4% 的情况下，老年抚养比超过 14%。

## 二. 二胎政策

在模型的基本假设中，我们假设一个家庭只生一个孩子，称为“方案一”。现在我们讨论“方案二”——“二胎政策”，即“夫妻双方均为独生子女可生育两个孩子”这一政策将给我国人口造成的影响。

为了更清楚地阐述这一影响，我们引入总和生育率的概念。我们可以把年龄别生育率看作一个妇女在该年龄时平均生的孩子，于是各个不同年龄段的生育率分布可以看成是一个妇女处在不同年龄段生育孩子数的分布。我们把这一分布称为生育模式。而总和生育率等于每个妇女一生中一共生育的孩子数。

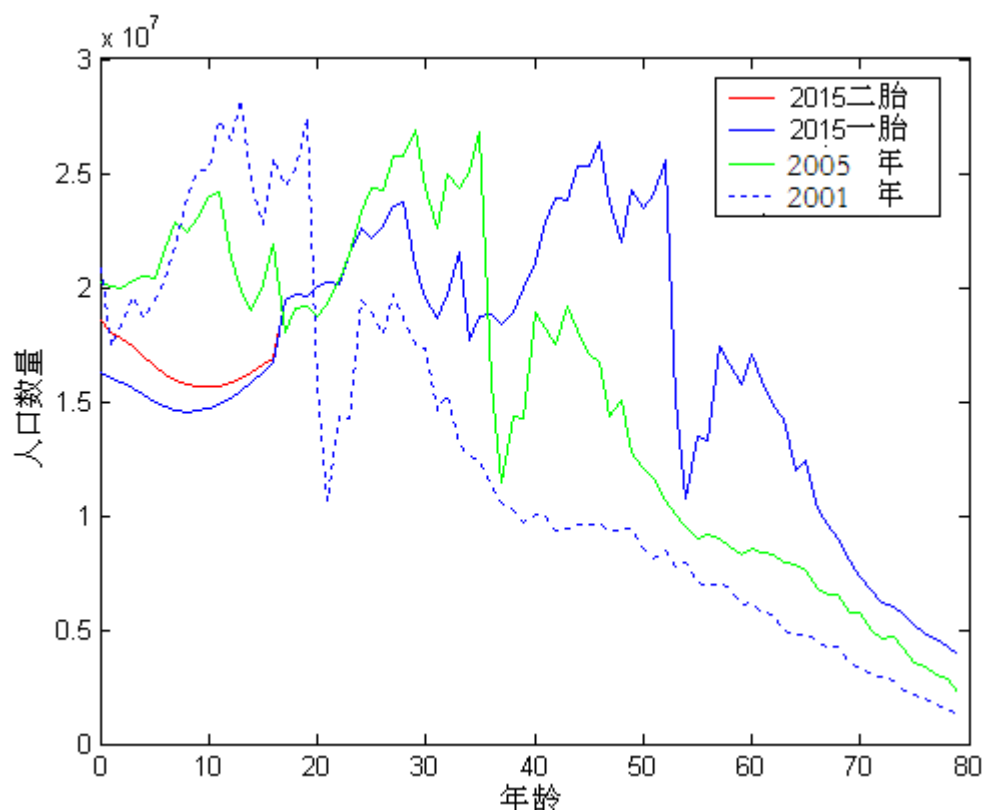
我国农村妇女的总和生育率在计划生育实行后一直维持在 1.5 以上，即使在 2000 年后，农村独生子女的比重很小，而夫妻均为独生子女的概率更小，于是可以认为该政策对农村的影响可以忽略，农村生育模式不变。而城市的总和生育率自 2001 年起非常接近 1，我们认为所有的 2001 年以后的城市出生的都是独生子女，于是方案二对城市的影响很大。下面是定量分析。考查 2001 年我国农村分胎次的生育率随年龄的分布：





农村每个妇女平均有 0.457 个二胎，而城市平均只有 0.146 个，但是它们的相对分布确实非常相近的。事实上，从图中可以看出，把城市妇女的第二胎的生育年龄结构按比例（0.457/0.146）放大后和农村妇女的相应的分布形状基本相同。这是容易理解的，由于受到生理条件的限制，在没有其他政策性引导的情况下（如规定生第二胎的最早年龄），妇女生二胎的年龄相对分布应基本稳定。另外，我们还发现，在 35 岁后生育的概率很小，于是从最早的独生子女开始生育到他们 35 岁这一时期（即大约 2001—2015），我国城市妇女的生育模式逐渐向二胎过渡。而她们的孩子这时还未到生育年龄，于是可以认为生育年龄分布向二胎变化，到 2015 年完全达到。也就是说，到 2015 年，城市妇女的生育年龄结构可看成第一胎和第二胎的叠加。我们假定第一胎的生育模式不变（其实观察上图可知，图形应该变尖锐一些，且峰将提前，但鉴于变化过于复杂，本文不讨论）。对于二胎，我们把 2001 年农村妇女的二胎的年龄分布按比例放大到总和生育率为 1。这样我们就得到了二胎完全普及的 2015 年中国城市的生育模式。从 2001 年到 2015 年，我们考虑最简单的情况，即假定变化是线性的，可得各年的生育率，即  $b_i^{urban}(k)$ 。同时，我们假定  $d_i(k)$  和城乡比不变。事实上，从所给的现有数据来看，城乡比变化缓慢。而且，在 2001–2015 年这段时间内，城市和农村的总和生育率比较接近，平均都在 1.5 左右，所以城乡比有几个百分点的差别并不会太影响结果。

用上述方法得到的 2015 年的人口分布如图。



是否实行“二胎政策”对我国 2015 年人口的影响图

观察上图，我们发现，从 2001 年到 2015 年我国人口的年龄分布变化是非常剧烈的。但总的来说，曲线趋于平坦，也就是说，人口结构慢慢由均匀化的趋势。此方案实际上是对城镇中双方都是独生子女父母家庭的一种补偿。使城乡居民在生育政策上趋于平等。减缓所谓的“四二一”家庭的压力和人口老化的压力。减缓所谓“城乡人



口素质倒挂”问题。但是，世界上不可能有十全十美的公共政策。上述调整同样有现在看来是“负面”的效应：将使我国 2015 年时 0-10 岁的人口增加 10%左右；人口总数在 2015 年时比“方案 1”增加 1 亿左右。

## 模型灵敏度分析

对多元线性回归模型，如下式：

$$Y = 164551.99 + 942.172X_1 + 119.794X_2 - 6610.520X_3 - 25.875X_4$$

进行灵敏度分析，得到表 2 结果：

表 2

全部自变量	偏回归系数	标准误差	标准化偏回归系数	T 值	P 值
截距	164551.99	22543.187		7.299	0.000
$X_1$	942.172	1760445	0.986	5.340	0.002
$X_2$	119.794	414.954	0.072	0.289	0.783
$X_3$	-6610.520	3783.623	-0.124	-1.747	0.131
$X_4$	-25.875	145.111	-0.01	-0.178	0.864

由表 2 第 2 列，自变量  $X_1$ ， $X_2$  对因变量 Y 是正相关，自变量  $X_3$ ， $X_4$  对因变量 Y 是负相关。由表 2 第 4 列的标准化偏回归系数可知，各自变量对人口增长数量的影响程度依次是：时间顺序>死亡率>出生率>性别比率，很明显，多元线性回归方程可以对人口数量进行预测，我国近 5 年的人口政策基本不变，可以从表 1 计算 2001—2005 年的平均出生率，死亡率，性别比率，他们分别是 12.668, 6.434 和 106.18，2006 和 2007 年时间顺序分别取 6 和 7，将这些数据带入多元线性回归方程，对我国 2006 和 2007 年人口数量预测值分别为 132093.0062 和 133035.1782 万人。

从多因素分析结果看，相比较而言，影响人口增长的主要因素依次是：时间顺序，死亡率，出生率，性别比率。从表 1 的实际效果看，历年时间顺序，出生率是增加人口数量的重要因素。就出生率而言，从平均意义讲逐年下降，从 2001—2005 年有趋向 12.2 与 12.6 之间达到平衡，出生率一旦超出这个平衡区间，对人口增长起负作用，死亡率升高。从表 1 实际效果上看，性别比率基本上在 100.4 至 100.63 之间，这也许是男性人口与女性人口之间的差基本保持不变，造成性别比率变动非常微小，对人口减少作用不大，从而随死亡率增多或减少，对人口数量的减少或增加起重要作用的因素。从统计学角度影响人口增长的关键因素是时间顺序、死亡率和出生率，这些为有关制定人口政策部门的领导找出解决关键因素的方法提供帮助。

由表 1 方程的绝对误差和相对误差均非常小，可以对人口数量作近期预测，用多元线性回归方程对我国 2006 年和 2007 年人口数量进行预测和表 1 数据总结，发现在目前人口政策不变的情况下，人口近期每平均增长在 850 至 1100 万人左右，可以预见相隔



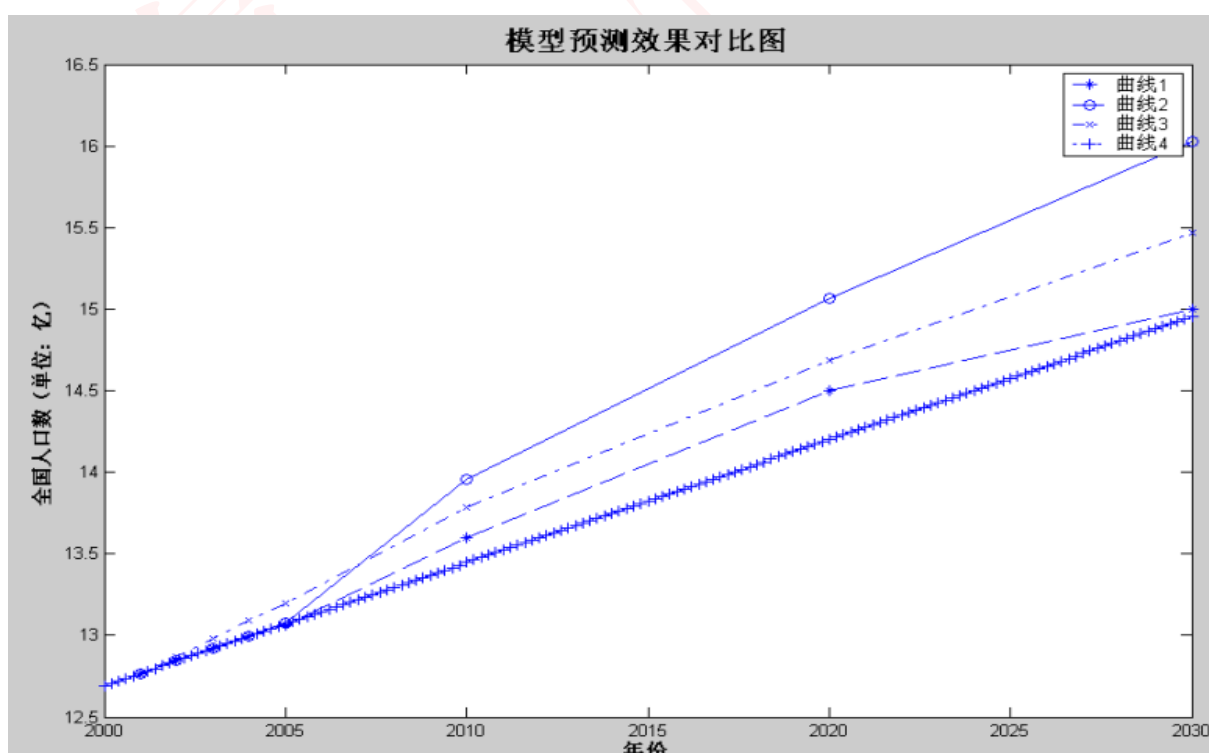
12 年至 13 年后中国人口数量将增加 1 个亿。

出生性别比偏高会造成严重的后果, 首先将会提高人口更替水平, 其次将使未来出生人数减少, 造成较大的人口数量波动, 再次将降低人口身体素质, 使人口老化速度加剧。因此, 要采取必要措施, 我们建议: 可以提高农村的社会保障水平, 改造小农经济, 加强性别鉴定和人工流产的法律管制, 完善和强化相关法律等。这些措施在一定程度上能起到一些作用, 但要真正解决我国婴儿出生性别比偏高问题, 还需要各个有关学科和政府部门的大力合作和全社会的共同努力才能完成。

## 模型评价

### 一. 模型对比与分析:

	(亿人)	2001	2002	2003	2004	2005	2010	2020	2030
1	官方参考值	12.76 27	12.84 53	12.92 27	12.998 8	13.0628	13.6	14.5	15
2	多元线性回归模型	12.76 27	12.84 53	12.92 27	12.998 8	13.0756	13.9567	15.0689	16.0245
3	L 模型	12.76 27	12.87 04	12.97 90	13.088 5	13.1990	13.7869	14.6849	15.4697
4	一元线性模型	12.76 27	12.84 31	12.91 85	12.993 8	13.0692	13.4461	14.1998	14.9535
5	Logistic 模型	12.76 27	12.87 53	12.98 91	13.103 8	13.2195	13.8135	15.0829	16.4689





由以上图，表可知，Logistic 模型预测的误差较大，原因在于时间较长，人口数据变化大，必然  $r$ ,  $x_m$  值变化大，因此误差较大且不稳定。

一元线性模型预测的精度最差，原因在于它考虑的影响因素少，导致人口数据随时间变化大。

多元线性回归模型在短期内精度最好，但对中长期外推预测，由于置信空间在扩大，误差较大，尤其在转折时期，函数形式在发生变化，误差更大。

Leslie 模型预测在长期预测的精度好，精度介于 Logistic 和多元线性模型之间。事实上，当矩阵  $\bar{L}$  不变的情况下，该模型有稳定性。

下面说明时间充分长后（即  $k \rightarrow \infty$ ）人口的年龄结构及数量的变化。根据  $b_i$  和  $s_i$  的定义，矩阵  $\bar{L}$  中的元素满足

$$0 < s_i \leq 1, \quad i=1, 2, \dots, n-1 \quad (7)$$

$$b_i \geq 0, \quad i=1, 2, \dots, n, \quad \text{且至少有一个 } b_i > 0 \quad (8)$$

容易看出， $\bar{x}(k)$  的稳定性态完全由  $L$  矩阵决定。关于  $L$  矩阵，我们有如下两个定理。

定理 1：L 矩阵有唯一的正特征根  $\lambda_1$ ，且它是单重的， $\lambda_1$  对应的正特征向量

$$\bar{x}^* = \left[ 1, \frac{s_1}{\lambda_1}, \frac{s_1 s_2}{\lambda_1^2}, \dots, \frac{s_1 s_2 \dots s_{n-1}}{\lambda_1^{n-1}} \right]^T \quad (9)$$

L 矩阵的其他  $n-1$  个特征根  $\lambda_k$  都满足

$$|\lambda_k| \leq \lambda_1, \quad k=2, 3, \dots, n \quad (10)$$

这定理表明 L 矩阵的特征方程

$$\lambda^n - (b_1 \lambda^{n-1} + b_2 s_1 \lambda^{n-2} + \dots + b_{n-1} s_1 s_2 \dots s_{n-2} \lambda + b_n s_1 s_2 \dots s_{n-1}) = 0 \quad (11)$$

只有一个单正根  $\lambda_1$ ，且容易验证  $\bar{L}\bar{x}^* = \lambda_1 \bar{x}^*$ 。

定理 2：若 L 矩阵第一行有两项顺次的元素  $b_i$ ,  $b_{i+1}$  都大于零，则(10)式中仅不等号成立，即

$$|\lambda_k| < \lambda_1, \quad k=2, 3, \dots, n \quad (12)$$

且由(6)式表示的  $\bar{x}(k)$  满足

$$\lim_{k \rightarrow \infty} \frac{\bar{x}(k)}{\lambda_1^k} = c \bar{x}^* \quad (13)$$



其中  $c$  是由  $b_i$ ,  $s_i$  及  $x(0)$  决定的常数。

从上述定理可以对时间充分长 ( $k$  充分大) 后人口按年龄组的分布  $\bar{x}(k)$  的性态作出如下分析。(为方便起见, 记  $\lambda_1$  为  $\lambda$ )

1) 由 (13) 式直接有

$$\bar{x}(k) \approx c\lambda^k \bar{x}^* \quad (14)$$

这表明  $k$  充分大时, 人口按年龄组的分布  $\bar{x}(k)$  趋向稳定, 其各年龄组的数量占总数量的比例, 与特征向量  $\bar{x}^*$  中对应分量占总量的比例是一样的, 即  $\bar{x}^*$  就表示了人口按年龄组的分布状况, 故  $\bar{x}^*$  可称为稳定分布, 它与初始分布  $x(0)$  无关。

2) 由 (14) 式又可得到

$$\bar{x}(k+1) \approx \lambda \bar{x}(k)^{**}$$

或者更清楚地写作

$$x_i(k+1) \approx \lambda x_i(k) \quad i=1, 2, 3, \dots, n \quad (15)$$

这表明  $k$  充分大时, 种群的增长也趋向稳定, 其各年龄组的数量都是上一时间段同年龄组数量的  $\lambda$  倍, 即人口的增长, 完全由  $L$  矩阵的唯一正特征根决定。显然, 当  $\lambda > 1$  时, 人口数量递增; 当  $\lambda < 1$  时, 人口数量递减, 为固有增长率。

3)  $\lambda = 1$  时, 人口总量不变。由 (11) 式可知, 这个条件等价于

$$b_1 + b_2 s_1 + \dots + b_n s_1 s_2 \dots s_{n-1} = 1 \quad (16)$$

若将 (16) 式左端记作

$$R = b_1 + b_2 s_1 + \dots + b_n s_1 s_2 \dots s_{n-1} \quad (17)$$

由这个表达式可以知道,  $R$  表示一个个体在存活期内繁殖的平均数量, 称为总和生育率, 显然,  $R=1$  时, 人口的总量不变。由 (9) 式, 可知此时稳定分布为

$$\bar{x}^* = [1, s_1, s_1 s_2, \dots, s_1 s_2 \dots s_{n-1}]^T \quad (18)$$

根据 (14) 式, 这个结果表明

$$x_{i+1}(k) \approx s_i x_i(k) \quad (19)$$

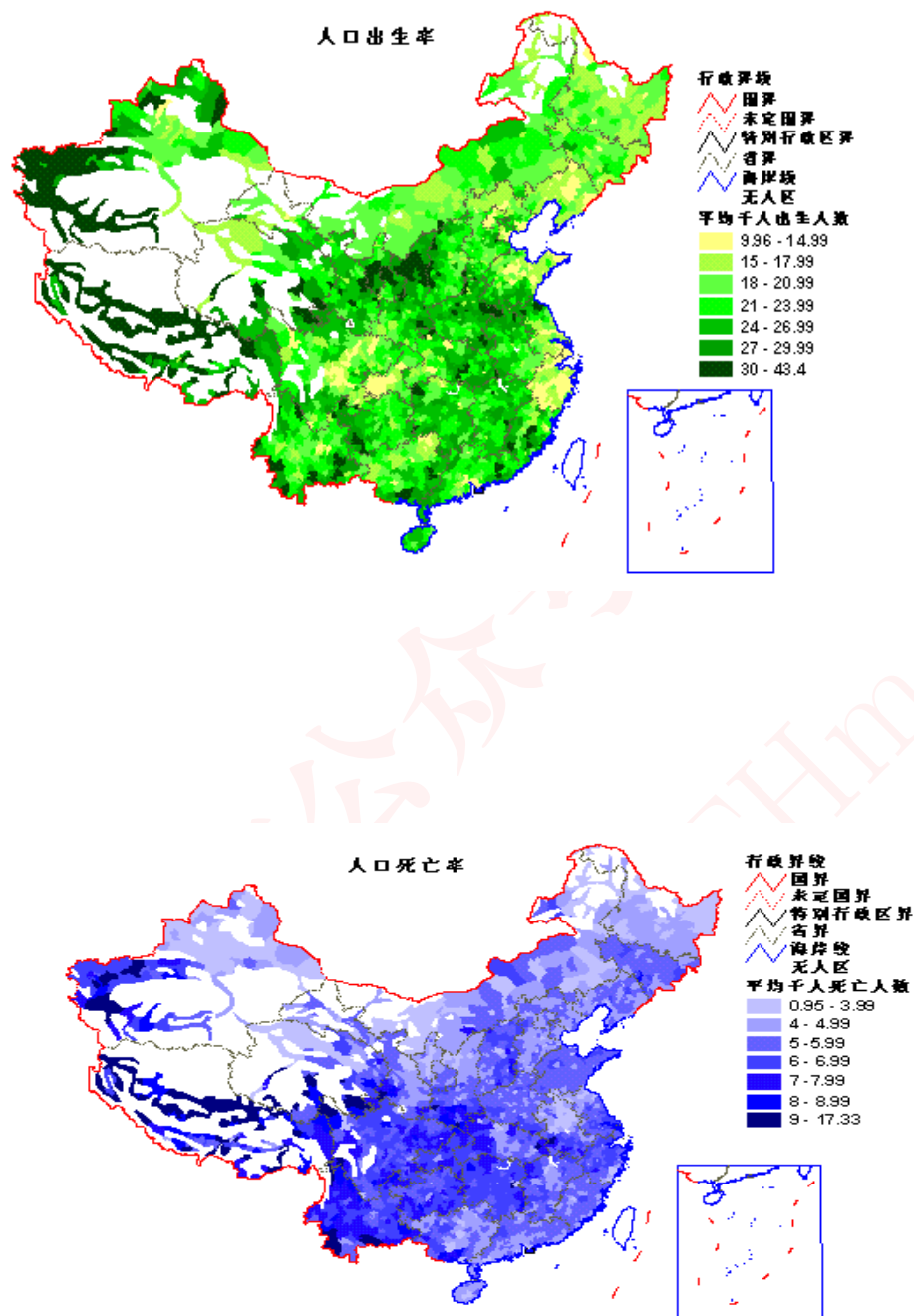
即当  $k$  充分大且  $R=1$  时, 存活率  $s_i$  等于相邻两年龄组的人口数量  $x_{i+1}(k)$  与  $x_i(k)$  之比。

## 二. 模型优缺点:

由于所给数据有限, 我们在建立模型时忽略了一些因素, 因此模型还存在一些缺点,



比如，我们没有考虑地区差异造成的影响。我们上网查了一些相关资料，如下为我国人口出生率和死亡率随地区的分布情况。



据近日公布的“五普”人口地区分布数据显示，我国东、中、西部地区人口密度分别为 452.3 人/平方公里、262.2 人/平方公里和 51.3 人/平方公里；东部人口密度是西部的 8.8 倍。自 30 年代著名人口地理学家胡焕庸先生发现，并提出中国人口分布的胡



换庸线以来，我国人口分布的大格局不仅未发生变化，且东多西少分布格局更加显著。

1990 年“四普”数据统计显示，我国东部地区（11 个省、市）人口密度为 392.0 人/平方公里，中部地区（8 个省）为 241.4 人/平方公里，西部地区（12 个省、区、市）为 52.5 人/平方公里。与此相比，“五普”数据显示，我国人口在东、中、西部地区分布情况，由多至少为东部 4.9 亿人、中部 4.2 亿人、西部 3.6 亿人；而东、中、西部地区面积大小排序恰好倒置，依次为 108.6 万平方公里、158.5 万平方公里和 692.7 万平方公里；人口密度分别为 452.3 人/平方公里、262.2 人/平方公里和 51.3 人/平方公里。由此可见，我国人口分布东多西少的格局没有发生明显变化，且日渐显著。

据有关专家分析，我国人口东多西少分布格局的形成，一是与客观自然地理状况相关，二是与社会经济发展密不可分。主要是 70 年代末以来，改革开放和实施沿海发展战略，以至西部地区形成人口净迁出现象的出现。新世纪，“孔雀东南飞”、“一江春水向东流”的人口东移现象会不会有所改变？西部大开发战略的实施，将会对我国人口分布宏观格局产生哪些影响？目前还都不得而知。

另外，模型一对于长期的预测比较准确，比较接近实际数据。模型二对于中短期的预测比较准确。

## 参考文献

- 【1】 姜起源 谢金星 叶俊，数学建模（第三版），北京市：高等教育出版社，2006 年
- 【2】 工程数学学报编辑委员会，2006 年工程数学学报，西安：西安交通大学出版社，2006 年
- 【3】 徐士良，数值分析与算法，北京市：机械工业出版社，2003 年
- 【4】 于景元，郭宝珠，朱广田，人口分布参数系统控制理论，华中理工大学出版社，1999 年
- 【5】 杨书章，王广州，孩次性别递进比研究，人口研究，2006 年
- 【6】 国家教委高效人口学研究及培调办公室，人口科学的探究与开拓，北京：北京大学出版社，1997 年
- 【7】 新华网湖北频道，中国人口出生率和死亡率  
[http://www.hb.xinhuanet.com/zhuant/2004-09/05/content\\_2809171.htm](http://www.hb.xinhuanet.com/zhuant/2004-09/05/content_2809171.htm)  
2007 年 9 月 22 日

