第24卷 增刊2
2007年12月

工 程 数 学 学 报
CHINESE JOURNAL OF ENGINEERING MATHEMATICS

Vol. 24 Supp.2
Dec. 2007

文章编号:1005-3085(2007)08-0047-07

中国人口区域结构向量模型

聂 磊， 陈 昆， 杨 慧
指导教师： 刘晓石
(四川大学，成都 610065)

编者按： 本文通过数据分析，找出了影响我国人口增长的主要因素，在此基础上建立合理的人口预测模型，对其中的参数拟合有特色，模型计算方法正确，结论可信，是一篇较好的数学建模参赛论文。

摘 要： 本文研究了影响人口的因素与我国人口变化之间的相互关系，对未来我国的人口增长趋势进行分析、预测。首先考虑区域和性别的差异，将我国人口的相关信息整合为由城市男性、城市女性、城镇男性、城镇女性、乡村男性、乡村女性组成的相关矩阵。在此基础上结合影响我国人口增长的，人口密度、生育率、死亡率、迁移率以及人口性别比等因素。综合这些因素并结合经典人口发展偏微分方程，通过完善和改进，建立了基于人口区域结构向量的一阶偏微分方程组模型。为了求解总模型中的三个未知向量分布函数，我们分别建立三个子模型：首先，对生育率数据进行曲线拟合建立了基于高斯分布的生育率分布函数子模型。其次，使用了指数分布和 Weibull 分布建立了死亡率分段函数子模型。再次，建立了理想化迁移率子模型。最后对方程进行离散化处理，运用计算机模拟方法求解。

关键词： 人口发展偏微分方程；人口区域结构向量模型；生存分析；计算机模拟

分类号： AMS(2000) 91D10 **中图分类号：** O29 **文献标识码：** A

1 问题的重述与社会背景

作为一个人口大国，一直以来我国的发展在一定程度上始终受到人口问题的制约。如今我国正在经历全面建设小康社会的快速转型期，人口发展面临着前所未有的复杂局面。为了全面落实科学发展观、实现低生育水平、提高人口素质、改善人口结构、引导人口合理分布、保障人口安全、人口与经济社会资源环境的协调和可持续发展等目标[1]。现我们根据《国家人口发展战略研究报告》提供的相关分析以及《中国人口统计年鉴》中的部分数据，着重探讨了以下两个问题：

1) 分析给出的人口数据，找出它们之间的相互关系，建立中国人口增长的数学模型；
2) 利用建立起来的模型，对我国人口短期以及中长期的增长趋势进行预测。

2 基本假设

1) 假设除少数异常点外其余所给数据真实；
2) 假设所给数据中不包括港、澳、台的人口信息；
3) 假设人口只受我国国内的出生率、死亡率和迁移因素影响，不考虑国家间的移民；
4) 假设人口在城、镇、乡之间迁移后，它的生育率与死亡率函数立即发生相应改变；
5) 假设国内社会环境基本稳定，没有影响人口发展的重大军事事件和严重的流行性疾病以及自然灾害等。

3 符号说明

r：表示人的年龄　r_m：表示人的最大年龄　t：表示时间

以下所有角标当中 $i=1$，2，3，4，5，6

1代表城市男性(以下简记城男)，2代表城市女性(以下简记城女)，

3代表城镇男性(以下简记镇男)，4代表城镇女性(以下简记镇女)，

5代表乡村男性(以下简记乡男)，6代表乡村女性(以下简记乡女)，

例如

$$P(r,t)=(p_1(r,t),p_2(r,t),p_3(r,t),p_4(r,t),p_5(r,t),p_6(r,t))^T.$$

$p_i(r,t)$：分别表示 t 时刻年龄为 r 的人口密度

$bir_i(r,t)$：分别表示 t 时刻年龄为 r 的生育率

$dea_i(r,t)$：分别表示 t 时刻年龄为 r 的死亡率

$k_i(t)$：分别表示表示 t 时刻的新生婴儿死亡率

$tra_i(r,t)$：分别表示 t 时刻年龄为 r 的迁出率

A：出生人口性别分配矩阵

$P(r,t)$：人口密度向量函数

$\mathrm{Bir}(r,t)$：生育率向量函数

$\mathrm{Dea}(r,t)$：死亡率向量函数

$K(t)$：新生婴儿死亡率向量函数

$\mathrm{Tra}(r,t)$：迁移率向量函数

$*$：表示两个向量的对应元素相乘

例如

$$A=(a_1,a_2,a_3)^T,\qquad B=(b_1,b_2,b_3)^T,$$

则

$$A*B=(a_1b_1,a_2b_2,a_3b_3)^T.$$

4 模型的分析和建立

4.1 总体模型的分析与建立

首先考虑到我国城、镇、乡之间的生产生活水平存在较大差异，从而导致各个区域间的生育率、死亡率等人口信息存在差异性，另外受我国历史与传统影响，各区域内部的人口信息又存在男、女性别之间的差异性，为了避免不同区域、不同性别造成的影响，使模型的预测结果更为准确且符合我国实际情况，故将我国人口的各方面信息划分为城男、城女、镇男、镇女、乡男、乡女六部分。

其次经过仔细分析和结合相关理论资料，可以得到影响我国人口增长主要由于：人口密度、生育率、死亡率、人口数量、迁移率以及出生人口性别比因素这六方面因素。

综合以上分析，在已有的人口发展偏微分方程[2]的基础上，可得到在时刻 t，年龄在 $(r,r+\Delta r)$ 内的人数为 $P(r,t)$。经过时间 Δt 后，这些人的年龄在区间 $(r+\Delta r,r+\Delta r+\Delta t)$ 内，人数变为 $P(r+\Delta t,t+\Delta t)\Delta r$。并且在 t 时刻年龄为 r 的人数的变化只由在 Δt 时刻内年龄为 r 的死亡人数和迁移人数决定，并令 $\Delta r\to 0$，$\Delta t\to 0$，并结合边界条件，得到人口区域结构向量模型

$$\begin{cases}\frac{\partial P}{\partial t}+\frac{\partial P}{\partial r}=-[Tra(r,t)+Dea(r,t)]*P(r,t), & \\ P(0,t)=(1-K)*A\int_0^{r_m}Bir(r,t)*P(r,t)\mathrm{d}r, & \\ P(r,0)=P_0(r), & t>0,\ 0<r<r_m, \\ P(r,t)=0, & t>r_m.\end{cases}\tag{1}$$

正常情况下，出生性别比是由生物学规律决定的，但是不同区域也呈现一定差异，保持在103~107之间，本文根据历史数据计算出平均出生人口性别，为了方便向量计算引入出生人口性别分配矩阵 A

$$A=\begin{pmatrix} 0 & 0.528 & 0 & 0 & 0 & 0 \\ 0 & 0.472 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.541 & 0 & 0 \\ 0 & 0 & 0 & 0.459 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.545 \\ 0 & 0 & 0 & 0 & 0 & 0.455 \end{pmatrix}$$

由于总体模型中生育率、死亡率、迁移率的向量函数未知，为了求解此模型，通过往年的数据分别建立生育率、死亡率、迁移率的经验模型。

4.2 生育率子模型的建立

由《年鉴》中的数据可以得到从2001到2005年的城、镇、乡的妇女生育率，运用 matlab 曲线拟合工具箱进行回归分析与数据拟合。综合分析多种拟合方式的试验效果后，得到高斯函数的拟合效果最为理想。故认为年龄为 r 的妇女的生育率函数近似为高斯函数

$$bir_j = a_j e^{-((r-b_j)/c_j)^2}, \tag{2}$$

其中 a_j，b_j，c_j 作为参数，$j=1, 2, 3$，分别代表城、镇、乡。

求解参数 a_j，b_j，c_j 对提供的数据进行曲线参数拟合。拟合过程中对年龄进行归一化处理，即 r =实际年龄/100，例如30岁的妇女年龄为 $r=0.3$。最后得到2001年、2002年、2004年、2005年的参数拟合值(见表1)。由于《年鉴》所给数据在运用计算机进行回归时发现数据异常，故排除2003年的生育率数据。

表1: 生育率参数拟合值

区域	a	a的置信区间	b	b的置信区间	c	c的置信区间	R_2
城01	127.4	(119.3, 135.6)	0.2595	(0.2573, 0.2617)	0.0412	(0.0389, 0.0451)	0.9695
镇01	150.4	(140.4, 160.5)	0.2515	(0.2492, 0.2538)	0.0415	(0.0331, 0.0447)	0.9663
乡01	182.3	(163.6, 201)	0.2464	(0.2425, 0.2502)	0.0462	(0.0407, 0.0517)	0.9194
城02	113.5	(107.9, 119.2)	0.2598	(0.2579, 0.2617)	0.0462	(0.0435, 0.0488)	0.980
镇02	148.4	(136, 160.7)	0.2476	(0.2446, 0.2505)	0.0429	(0.0388, 0.0470)	0.9480
乡02	177.2	(156.2, 198.2)	0.2465	(0.2417, 0.2512)	0.0492	(0.0425, 0.0560)	0.8923
城04	123.6	(117.5, 129.6)	0.2563	(0.2545, 0.2582)	0.0458	(0.0432, 0.0484)	0.9811
镇04	147.1	(135.7, 158.4)	0.2479	(0.2448, 0.2510)	0.0493	(0.0449, 0.0536)	0.9525
乡04	167.1	(149.6, 184.7)	0.2446	(0.2399, 0.2493)	0.0546	(0.0480, 0.0613)	0.9104
城05	95.85	(90.48,101.2)	0.2548	(0.2524,0.2572)	0.0527	(0.0493, 0.0561)	0.9736
镇05	134.7	(121.5, 148)	0.2412	(0.2372, 0.2452)	0.0499	(0.0443, 0.0556)	0.9226
乡05	154.9	(137.6, 172.2)	0.2433	(0.238, 0.2486)	0.0580	(0.0504, 0.0656)	0.8965

对上面表格中的数据进行观察、分析，可以得到：从2001到2005年(不包括2003年)城、镇、乡的生育率函数中只有参数 a 在进行有规律性的变化，参数 b 和 c 基本不变。为求 a 的

规律，运用 SPSS 软件进行倒数曲线模型拟合后，可求得 a 随时间的变化方程如下

$$a_1 = 90.575 + 38.25/t, \quad R^2 = 0.956, \tag{3}$$

$$a_2 = 134.283 + 18.031/t, \quad R^2 = 0.727, \tag{4}$$

$$a_3 = 153.448 + 31.796/t, \quad R^2 = 0.778. \tag{5}$$

由于根据这个倒数曲线模型求得 R^2 值都比较接近于1，从而说明 a 的方程与实际情况相符合，即随着我国社会经济的不断发展，人民生活状况的日益改善，我国城、镇、乡的生育率会逐渐降低并趋于一个稳定值。由于从2001到2005年(不包括2003年)城、镇、乡的生育率函数中参数 b 和 c 基本不变，所以对参数 b 和 c 取平均值；最后根据以上参数结果，利用常数变易法得到城、镇、乡在 t 时刻年龄为 r 的生育率函数，即有生育率模型

$$Bir(r,t) = \begin{pmatrix} 0 \\ (90.575 + 38.250/t)\mathrm{e}^{-((r-0.2576)/0.04665)^2} \\ 0 \\ (134.283 + 18.301/t)\mathrm{e}^{-((r-0.468)/0.04448)^2} \\ 0 \\ (153.448 + 31.796/t)\mathrm{e}^{-((r-0.2464)/0.0462)^2} \end{pmatrix}. \tag{6}$$

本文利用数据对总体的分布形态作出推断，假设生育率服从高斯分布，根据样本值落在总体各个区间的频数与总体落在该区间的期望频数的差值构造服从卡方分布的 Pearson 统计量[3]，并利用 SPSS 软件得到卡方统计量的观测值，依据卡方分布表计算出观测值对应的概率 p 值为0.137。该 p 值大于显著性水平0.05，故认为样本服从高斯分布具有统计意义。

4.3 死亡率子模型的建立

以往研究的寿命分布，从本质上说是围绕失效率考虑的。

设随机变量 ξ 表示人的寿命，其分布函数为 $F(r)$ 表示寿命不超过 r 的概率。设 ξ 的分布密度为 $f(r)$，记寿命大于 r 的概率为 $R(r)$

$$R(r) = P(\xi > r) = 1 - F(r), \tag{7}$$

当某人在时刻 r 存活，但他在区间 $(r, r+\Delta r)$ 内死亡的概率为

$$P(\xi \le r + \Delta r | \xi > r) = \frac{F(r+\Delta r) - F(r)}{1 - F(r)} \approx \frac{f(r)\Delta r}{R(r)}, \tag{8}$$

并且考虑社会比较稳定和时间较短的情况下，死亡率与时间几乎无关，所以年龄为 r 的人的死亡率为

$$dea(r,t) = \frac{f(r)}{R(r)}. \tag{9}$$

1) 在青壮年期，人的死亡不是由人体自身的衰老所引起，而是由外界的大量随机因素引起，因此无记忆效应，因此青壮年时期的寿命分布服从指数分布[4]，由此可得该段死亡率为

$$dea(r,t) = \lambda, \tag{10}$$

为确保准确性，通过对青壮年期的死亡率与时间 t 的相关分析，分别采用了三个相关系数：Pearson 相关系数，Spearman 相关系数和 Kendall 相关系数对模型进行检验(见表2)。分

析表2数据，由于相关系数较低并且有显著意义，可以得出结论：青壮年时期的死亡率与时间几乎没有相关性，故求平均值即可。

表2: 城、镇、乡的青壮年期死亡率与时间 t 的相关性

	Pearson	p值	Spearman	p值	Kendall	p值
城男	0.096	0.251	0.171	0.004	0.232	0.005
城女	0.004	0.065	0.050	0.399	0.060	0.472
镇男	0.173	0.038	0.148	0.012	0.204	0.014
镇女	0.135	0.105	0.082	0.182	0.109	0.190
乡男	0.378	0.000	0.268	0.000	0.384	0.000
乡女	0.166	0.046	0.210	0.000	0.266	0.001

2) 在衰老期，由于人的死亡率主要由人体自身的衰老引起，所以它近似服从 Weibull 分布[5]，即有衰老期的死亡率(见公式(11))，根据公式(11)，结合相关数据运用 SPSS 软件进行曲线参数拟合，得到城男、城女、镇男、镇女、乡男、乡女的参数拟合值(见表3)

$$dea(r,t) = \lambda\alpha(\lambda r)^{\alpha-1}, \tag{11}$$

其中 $\alpha, \lambda > 0$ 为参数。

表3: 老期死亡率参数拟合值

	λ	λ 的95%置信区间	α	α 的95%置信区间	R^2
城男	1.559	(1.515, 1.603)	8.793	(8.177, 9.408)	0.861
城女	1.350	(1.323, 1.377)	13.265	(12.174, 14.356)	0.838
镇男	1.598	(1.507, 1.689)	8.728	(7.564, 9.891)	0.631
镇女	1.493	(1.451, 1.535)	9.289	(8.549, 10.028)	0.832
乡男	1.624	(1.591, 1.656)	8.740	(8.343, 9.136)	0.936
乡女	1.535	(1.510, 1.560)	9.189	(8.800, 9.578)	0.945

3) 新生婴儿死亡率的计算

由提供数据分析，观察可得新生婴儿死亡率与时间不相关，故分别求得城男、城女、镇男、镇女、乡男、乡女的新生婴儿死亡率的平均值。

由以上综合后得到了我国城男、城女、镇男、镇女、乡男、乡女的死亡率分段函数(见表4)。

表4: 死亡率分段函数

区域	新生婴儿死亡率	青壮年时期死亡率	衰老期死亡率
城男	6.1160	0.4479	$dea_1(r,t)=13.71(1.559r)^{7.793}$
城女	6.3440	0.2983	$dea_2(r,t)=17.91(1350r)^{12}$
镇男	6.4000	0.7566	$dea_3(r,t)=13.95(1.598r)^{7.728}$
镇女	10.9900	0.6132	$dea_4(r,t)=13.87(1.493r)^{8.289}$
乡男	19.3100	1.1561	$dea_5(r,t)=14.19(1.624)^{7.74}$
乡女	23.9360	0.7401	$dea_6(r,t)=14.11(1.535)^{8.189}$

4.4 迁移率子模型的建立

人口迁移和流动作为一种复杂的社会经济现象，对经济发展、人口分布起着重要影响。流动迁移率对生育率有着非常显著的影响，所以在进行人口预测时就必然要考虑迁移率。但由于人口迁移流动的概念界定和统计口径上的不同以及我国户籍制度等问题，加之各次普查和全国性的抽样调查得到的多是存量指标，故不能得到准确的人口迁移数据。所以按人口城镇化水平年均增长1个百分点计算。

5 模型求解与结果分析

5.1 计算机模拟求解

为方便计算，将时间离散化处理，变偏微分方程组(见公式(1))为差分方程组，用计算机进行模拟求解，算法如下：

1) 把2005年末的状态赋为初始状态，设置步长 s 为0.01年。

2) 通过递推方程，由 t 时刻的状态计算出 $t+s$ 时刻各个年龄段的状态。

3) 当时间 t 大于预测时间，则退出循环，输出结果，当时间 t 小于预测时间，则令 $t=t+s$，转到步骤2)。

5.2 口总数预测

短期预测方面，从2006年到2015年，我国人口总数净增长7000万，增长率达到5.3%，并且具有继续上涨趋势；中长期预测方面，我国人口总数在本世纪30年代左右达到峰值14.59亿，此后具有下降并趋于稳定的趋势。我国人口总数在2010年、在2020年、2030年、2040年分别达到13.59亿、14.44亿、14.52亿、14.14亿。

5.3 人口年龄结构预测

我国老龄化情况有逐年加剧的趋势，老年人口数量增多、老龄化速度快、高龄趋势明显。目前我国65岁以上老年人口已达1.078亿，占总人口的8.1%。到2016年左右，65岁以上老年人口将达到1.358亿人，比重增长到10%；到2027年左右，65岁以上的人口达到2.219亿，占总人口的15%。到2044年左右预计会形成老龄人口高峰平台，65岁以上老年人口达3.479亿人，比重突破25%；届时每4人中就有1名老年人。

5.4 城镇乡比例情况

我们城镇化水平在未来40年内将逐步提高，农村人口比例将缓慢减少，更多的乡村人口被城镇化。城镇化农村中人口中，大多数进入镇，少部分进入城，镇人口比例与城市人口比例相比较变化更为明显。

5.5 结果分析

将利用模型预测得到的人口总数结果与《国家人口发展战略研究报告》中的结果进行比较，发现短期内模型预测与报告结果几乎相同，但对于中长期预测，模型预测数据较报告提供数据略为偏小。经分析，主要原因是死亡率数据较少，无法得到死亡率与时间的相关规律，从而导致不能建立能较好符合实际的死亡率经验模型。

另外分析模型预测出的人口年龄结构和城镇乡人口比例与《报告》预测数据进行比较，发现二者基本相同，说明预测方法较为科学、预测准确率较高。

参考文献：

[1] 国家人口发展战略研究报告[OL]. www.chinapop.gov.cn，2007年9月23日
[2] 姜启源，谢金星，叶俊. 数学模型[M]. 北京：高等教育出版社，2003年
[3] 张红兵，贾来喜，李潞. SPSS宝典[M]. 北京：电子工业出版社，2007年
[4] 王林书. 几种新寿命分布类[J]. 应用数学学报，2004, 27(3): 397-406
[5] 费培之等. 数学模型实用教程[M]. 成都：四川大学出版社，2004年
[6] 何晓群，刘文卿. 应用回归分析[M]. 北京：中国人民大学出版社，2001年

Vector Model for the Regional Structure of China's Population

NIE Lei, CHEN Kun, YANG Hui
Advisor: LIU Xiao-shi

(Sichuan University, Chengdu 610065)

Abstract: This paper studies the relationship between the factors that has an impact on the demographics and the changes of China's population. It can be used for analysis and prediction the trends of population growth in the future. Considering the regional and sexual differences, the relevant information of China's population is composed of urban males, urban females, town males, town females, rural males and rural females, all of which form a matrix. This division brings the population density, fertility, mortality, migration rate, sex ratio and other factors, which may affect China's population growth, into our consideration. Combining these factors with the classical partial differential equations for population development, we can establish the first order partial differential equation model. In order to solve three unknown vector distribution functions in the model, we set up three sub-models. Firstly by using the data of fertility to do a curve fitting, we establish the sub-model of fertility as a distribution function which is based on Gaussian function. Secondly, we use the exponential and Weibull distribution to set up the sub-model of mortality as a sectional function. Thirdly we establish the ideal sub-model of the migration rate. Finally, the function is discretized and solved through the computer simulation.
Keywords: partial differential equations of population development; vector model of the regional structure of the population; survival analysis; computer simulation