

2016 年第十三届五一数学建模联赛

承 诺 书

我们仔细阅读了五一数学建模联赛的竞赛规则。

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）与本队以外的任何人（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它公开的资料（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，我们愿意承担由此引起的一切后果。

我们授权五一数学建模联赛赛组委会，可将我们的论文以任何形式进行公开展示（包括进行网上公示，在书籍、期刊和其他媒体进行正式或非正式发表等）。

我们参赛选择的题号为（从 A/B/C 中选择一项填写）： C

我们的报名参赛队号为： 6100

参赛组别（研究生或本科或专科）： 本科

所属学校（请填写完整的全名） 山东科技大学

参赛队员（打印并签名）： 1. 曲文涛

2. 唐 静

3. 郭宇哲

日期： 2016 年 5 月 3 日

获奖证书邮寄地址：山东省青岛市黄岛区山东科技大学 邮政编码： 266590

收件人姓名： 曲文涛 联系电话： 13156279902

2016 年第十三届五一数学建模联赛

编 号 专 用 页

竞赛评阅编号（由竞赛评委会评阅前进行编号）：

评阅记录

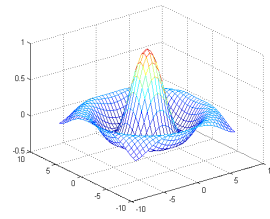
评 阅 人						
评 分						
备 注						

裁剪线_____裁剪线_____裁剪线_____

竞赛评阅编号（由竞赛评委会评阅前进行编号）：

参赛队伍的参赛号码：（请各参赛队提前填写好）：

2016 年第十三届五一数学建模联赛



题 目 “二孩政策”问题

摘 要

当前我国人口发展正面临着新的复杂局面,人口问题仍然是全面建设小康社会面临的重大问题。在过去一段时期,计划生育作为我国的基本国策,对中国人口的发展起到了积极的作用。因此,科学的预测人口结构,并采取相应的措施可以为经济发展提供重要的依据。本文主要研究不同政策下,人口结构变化趋势。

首先,我们根据搜集到的文献,统计出人口结构评价指标出现的频数,以此作为选择指标的参考。然后,本文建立了灰色关联度分析模型,得出非农业人口所占比与地区 *GDP* 有关,因此非农业人口所占比可以反映经济对人口结构的影响。结合搜集到的指标最终确定了出生率、死亡率、年龄结构、性别比例、人口密度、非农业人口所占比六个因素作为本文的评价指标。

其次,本文对附件中的数据进行整理,运用变异系数法得出性别比例、年龄结构(儿童组、青壮年组、老年组)、出生率、死亡率、非农业人口占比、人口密度的权值分别为 0.116、(0.062、0.032、0.047)、0.109、0.111、0.139、0.384,从而,建立出人口结构的综合评价模型。综合考虑经济、地理、人文等因素,选取了北京、广东、贵州、海南、辽宁、内蒙古、山东、四川、新疆维吾尔自治区、浙江十个省市作为评价的对象。对 A 地区的各项评价数据标准化后,得出 A 地区在 2014 年的综合得分为 14.57977,其人口结构的评价等级为 C。

再次,本文建立了人口年龄结构的房室模型,定义年龄组间人数传递比,构造了各年龄组的微分方程。通过对出生率变化进行类比拟合,得到出生率随时间变化的函数。本文利用黑龙江的各个年龄段的死亡人数进行验证传统的死亡率模型,从而得出各个年龄段的死亡率公式。在已知合理人口结构的基础上,将各参数带入房室模型得出最优出生率为 22.35%。

最后,本文引入了生产力水平、生育意愿、文化程度等因素修正了人口模型的系数和人口出生率,对上述的房室模型进行改进,从而预测出未来 20 年实行二孩政策以及全面放开生育政策的人口结构变化趋势。通过对比问题 3、4、5 预测的 A 地区二十年中的人口结构,得出结论:若 A 地区自由发展,则 20 年内会出现老龄化的趋势;若 A 地区全面实行二孩政策,则可以缓解 A 地区人口老龄化的趋势;在国民经济运行稳定的前提下,若 A 地区实现全面开放生育政策,则 20 年内可以较为明显的改善 A 地区的人口年龄结构。

此外,本文还对模型的优缺点进行了分析。

关键词: 房室模型 灰色模型 变异系数法 综合评价模型 微分方程

一、问题重述

1.1 问题背景

无论是对于我国目前的经济发展状况还是未来的远景规划，人口问题都是我国可以考虑一切问题的出发点和落脚点，因此人口问题的研究都具有十分重要的意义。为了处理好这一问题，我国多年来实施对多年来实施的严、紧计划生育政策对控制人口增长起到关键作用。在优生优育政策的指引下，我国人口质量显著提高，但也带来了不利影响，生育率偏低、男女比例失衡、人口老龄化情况严重等问题。

1.2 问题提出

自 2016 年，我国在全国范围内放开二孩政策。请查阅相关材料并参考附件材料，建立数学模型，解决如下问题。

(1) 建立人口结构可持续发展指标体系，相关指标应具有科学性、代表性、全面性。

(2) 建立人口结构可持续发展的评价模型。选取 10 个国内具有代表性的省市，对这些地区的人口结构进行评价分级。并结合所建的模型对 A 地区进行等级评价。

(3) 预测在 A 地区不实行二孩政策的情况下，未来 20 年的人口结构变化趋势；假设 A 地区实行二孩政策，给出该地区的最优出生率，使得该地区未来 20 年的人口结构更加合理。

(4) 二孩政策下，预测 A 地区未来 10-20 年按年龄划分的人口结构。

(5) 假如全面放开生育政策，在国民经济运行稳定的基础上预测 A 地区未来 20 年人口结构的变化趋势。

二、问题分析

2.1 任务一分析

针对问题一、二，要建立科学合理的指标，并对人口结构建立综合评价模型。因此，我们既要寻找合理建立评价指标的标准，又要考虑附件中评价 A 地区的各个指标。由于影响人口结构变化的因素较多，我们统计出搜集到的文献中人口结构评价指标出现的频数，以此作为选择指标的参考。为了使得指标更加全面，我

们分析非农业人口的比例一定程度上能反应经济的发展，并对两者进行关联度检验。确定指标之后，因为各个指标有按照年份的数据，因此在确定各个指标的权重时可以用变异系数法，相对层次分析法来说，减少了主观影响，更加客观。确定十个具有代表性的省市，需要从多个角度考虑，比如中国的位置——内陆或者沿海地区，经济发展程度，是否是直辖市，地形是否特殊等。

2.2 任务二分析

针对问题三，本文延续前文思路，人口结构要考虑出生率、死亡率、性别比例、年龄结构、人口密度、非农业人口占比等因素。但是非农业人口占比反应经济对结果的影响，因为各个地区的经济发展无法确定，所以此时不对非农业人口占比考虑。

对于出生率，由于 A 地区在 2010 年时发生了政策变化，所以使得出生率发生了突变，因此我们对于 A 地区只有五年可以应用的出生率数据，虽然每个地区的出生率不同，但是变化趋势是相同的，因此我们从其他省份的出生率变化中抽象出一个指数函数，进而对其进行拟合即可。

想要求出年龄结构的变化，就要知道各个年龄段的人口的死亡率，因此单纯的对 A 地区的年死亡率进行拟合是不合理的。我们搜集了死亡率的通用模型，并对黑龙江各个年龄段的人口死亡率拟合，对比这几种方法来确定死亡率的模型。性别比例在没有实行二胎政策时，其变化趋势会按照以前的趋势，因此建立灰色预测模型对其进行预测。

年龄结构包括三个年龄段，每个年龄段的比例会随时间变化而变化，用房室模型可以更清楚的分析出影响变化率的每个因素，因此可以得到每个年龄段人数变化的微分方程。解微分方程可以得到未来二十年中每一年各个年龄段的人数，把每个年龄段人数相加就可以得到总人数，相应的人口密度也就可以得到。

由以上对出生率死亡率个的分析，可以得出最终对人口结构影响的是年龄结构，因此开放二孩政策后，求最优出生率使得人口结构更加合理，也就是使得年龄结构更加合理。因此就是在建立合理的人口结构后，用房室模型反过来求出生率。

2.3 任务三分析

针对问题 4 和问题 5，我们考虑到即使开放二孩或者生育政策，也不会引起出生率的激增，因为人们的生育意愿的不同，以及地区之间文化程度的差异以及生产力的发展，种种因素都会影响人口结构的变化。因此我们引入这三个因素的影响因子，对房室模型进行优化，也对出生率进行了优化。基于改进的人口结构预测模型，对第四问进行求解。第五问中，全面开放了生育政策，因此出生率的系数会相应变大，且在国民经济运行稳定的前提下，因此可以将生产力的影响因素看做是 1，进而进行求解即可。

三、模型假设

基于对问题的分析我们做如下假设：

- ✚ 假设本文搜到的数据真实可靠；
- ✚ 假设所统计的年份中没有人口迁移所带来的人口数量的变化；
- ✚ 预测年限期间没有因大规模自然灾害而导致的死亡率的变化；
- ✚ 假设妇女的生育意愿和期望生育年龄在 20 年内不发生显著变化。则可以认为生育模式在这段时间内保持恒定；
- ✚ 假设题目所说的二孩政策为全面二孩政策；
- ✚ 假设全面放开生育政策后，国民经济运行稳定。

四、符号说明

本文所用主要符号变量符号及意义，此处只列出整篇论文中通用的符号，具体的符号会在引用时再加说明。

符号	解释
$\alpha(t)$	表示第 t 年的出生率
$\mu_i(t)$	表示第 t 年年龄为 i 岁人的死亡率
β_i	表示各年龄组间的传递比， $i=1,2$
b_i	表示各年龄组的死亡率， $i=1,2,3$
x_i	表示指标 A_i 的数据
W_i	表示指标 A_i 的权重
$x_i^{(0)}$	表示第 i 指标的参考数列
$y_i^{(1)}$	表示生成的均值数列
\hat{b}	表示预测方程的灰作用量

五、模型建立与求解

5.1 人口结构发展综合评价指标体系的建立

5.1.1 指标体系构建的原则

➤ 适合中国国情与评价对象的原则。

反映中国国情、明确评价对象是做好人口发展综合评价工作最优先考虑的问题。如果评价指标体系的选取偏离评价人口结构发展的内涵，那么即使数据再准确，所得出的结论也是不正确的。我们将搜集到的 70 多篇文献中所提到的指标进行计数处理，得到图 1。

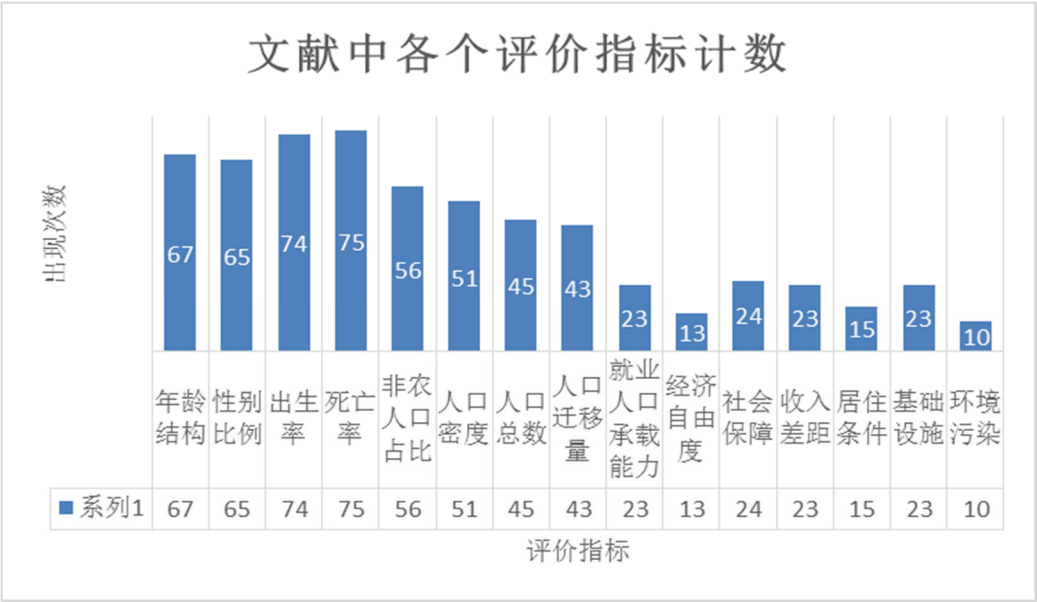


图 1 文献中各个评价指标计数

由图 1 可以直观的看出，性别比例、年龄结构、出生率和死亡率在人口结构的评价指标体系中的出现频率都比较高。

- 注重指标选取的准确性与系统性。
- 从准确性的角度来看：应正确选用总量指标与相对指标。总量指标是反映某种社会经济现象在一定条件下所达到的总规模、总水平的统计指标，结合附件所给数据，我们用人口密度来代替人口总数作为评价指标。
- 注重指标体系的科学性和可操作性。
- 全面、准确和系统的指标体系才能对评价对象进行科学的综合评价，因此我

们要评价一个对象，首先要有准确的数据，因此建立指标要结合附件数据。

➤ 注重指标的全面性和代表性

人口增长的影响因素中，经济因素是一个不可忽略的因素，因为城市和农村的人口增长是极不相同的，这主要是由于城乡经济发展不平衡导致，因此，我们假设某市的农业人口与非农业人口的比值与经济因素相关。分别选取了北京、新疆这两个地区，对非农业人口占比和地区总 GDP，人均 GDP，地区 GDP 指数这三个经济指标进行灰色关联度分析，对假设进行验证。

5.1.2 灰色关联度分析模型

1) 原理分析

➤ 确定分析序列

在对研究问题定性分析的基础上，确定一个因变量因素和多个自变量因素。设因变量数据构成参考序列 $X_i' (i=1,2,\dots,n)$ ，各自变量数据构成比较序列 $n+1$ 个数据序列成矩阵：

➤ 对变量序列进行无量纲化

一般情况下，原始变量序列具有不同的量纲或数量级，为了保证分析结果的可靠性，需要对变量序列进行无量纲化。

➤ 求差序列、最大差和最小差

➤ 计算关联系数

对绝对差值阵中数据作如下变换：

$$\xi_{0i}(k) = \frac{\Delta(\min) + \rho\Delta(\max)}{\Delta_{0i}(k) + \rho\Delta(\max)}.$$

➤ 计算关联度

比较序列 X_i 与参考序列 X_0 的关联程度是通过 N 个关联系数来反映的，求平均就可得到 X_i 与 X_0 的关联度

$$r_{0i} = \frac{1}{N} \sum_{k=1}^N \xi_{0i}(k).$$

➤ 依关联度排序

对各比较序列与参考序列的关联度从大到小排序，关联度越大，说明比较序列与参考序列变化的态势越一致。

➤ 模型应用

运用上述模型,我们得出北京市的非农业人口占比与三个经济因素之间的关联系数如表 1 所示:

表 1 北京市非农业人口占比与经济因素的关联系数

年份	$\xi_{01}(t)$	$\xi_{02}(t)$	$\xi_{03}(t)$
2010 年	0.336096809	0.399235818	0.999999907
2011 年	0.516032064	0.577460967	0.862172279
2012 年	0.870882905	0.865548055	0.838984953
2013 年	0.565766864	0.649789055	0.838984953
2014 年	0.398432329	0.483378515	0.817012167

最后分别对 $x_i(t)$ 与 $x_0(t)$ 的关联系数序列求算术平均可得下式,

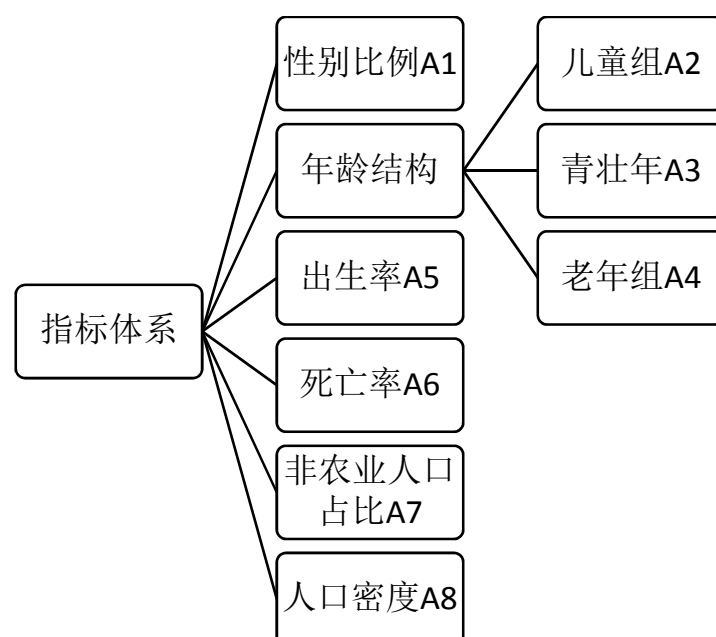
$$r_{01}=0.537, r_{02}=0.595, r_{03}=0.871.$$

同理,可以得到新疆的三个关联系数序列的算数平均值分别为, 0.579, 0.601, 0.836。因此可知,地区 GDP 指数与非农业人口占比的关系最大,其次是地区 GDP,再次是地区 GDP。

因此,非农业人口占比可以作为经济因素的替代指标。

5.1.3 人口结构发展综合评价指标体系的构建

从人口发展内涵出发,本文得出以下 6 个指标:



5.2 人口结构的加权综合评价模型

5.2.1 系数变异法确定指标权值

变异系数法是直接利用各项指标所包含的信息，通过计算得到指标的权重，是一种较为客观的赋值方法。在评价指标体系中，指标取值差异越大的指标，也就是越难以实现的指标，这样的指标更能反映被评价单位的差距。变异系数公式如下：

$$V_i = \frac{\sigma_i}{\bar{x}_i} (i = 1, 2, 3 \dots) \tag{1}$$

其中， V_i 表示第 i 项指标的变异系数、也称为标准差系数， σ_i 是第 i 项指标的标准差， \bar{x}_i 是第 i 项指标的标准差。

各项指标的权重为：

$$W_i = \frac{V_i}{\sum_{i=1}^n V_i} \tag{2}$$

对附件一中十个省份的数据，利用公式 (1) (2) 对其进行求权值，得出表 2：

表 2 人口结构可持续发展评价指标的权重

指标	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8
平均数	85.131	0.158	0.737	0.090	9.368	5.417	0.772	0.015
标准差	42.706	0.042	0.103	0.018	4.382	2.599	0.462	0.024
变异系数	0.502	0.266	0.139	0.203	0.468	0.480	0.599	1.654
权重系数	0.116	0.062	0.032	0.047	0.109	0.111	0.139	0.384

得出表 2 权重后，可以对各个指标建立综合评价模型。

5.2.2 对地区的人口结构进行综合评价

定义如下公式

$$y = x_i \cdot W_i. \tag{3}$$

其中， y 表示该地区的人口结构可持续发展评价得分； x_i 表示指标 A_i 的数据； W_i 表示指标 A_i 的权重。

根据 y 的数值我们将各地区人口结构可持续发展分为 A、B、C、D、E 五个等

级，对应关系如下：

表 3 人口结构可持续发展评价等级

等级	A	B	C	D	E
得分	≥ 15.5	≥ 15	≥ 14	≥ 13	< 13

5.2.3 分析模型建立与求解

本文中选取了新疆维吾尔自治区、内蒙古作为中国内陆的代表，北京作为直辖市的代表，山东作为东部地区的代表，辽宁是东北三省的代表，浙江、广东作为沿海经济较为发达地区的代表，海南作为海岛的代表，四川有特殊的地形、贵州则代表南方的一般内陆地区。

对上述十个地区的数据（详见附件一）进行处理，得到表 4

表 4 十个的评价地区得分

省份	2010 年	2011 年	2012 年	2013 年	2014 年
北京	12.30742	12.54738	12.68187	12.96523	12.35193
广东	14.82313	15.35612	15.12976	14.90702	15.82013
贵州	15.61716	15.74331	14.99064	15.13727	14.83733
海南	15.51645	15.22559	15.84011	15.61523	16.03894
辽宁	13.56291	13.42483	13.43476	13.54564	13.68404
内蒙古	14.58818	14.51429	14.53676	14.4966	14.52186
山东	14.14645	14.07118	14.32852	14.40803	14.61552
四川	14.16903	14.41402	15.06808	14.36261	13.72508
新疆	14.68768	14.6942	14.57817	14.83491	14.81449
浙江	13.80487	13.12303	14.22903	14.55073	14.89128

因此可以得出相应的等级评分如表 5 所示：

表 5 十个地区的等级

省份	2010 年	2011 年	2012 年	2013 年	2014 年
北京	E	E	E	E	E
广东	C	B	B	C	A
贵州	A	A	C	B	C
海南	A	B	A	A	A
辽宁	D	D	D	D	D
内蒙古	C	C	C	C	C
山东	C	C	C	C	C
四川	C	C	B	C	D
新疆维吾尔自治区	C	C	C	C	C
浙江	D	D	C	C	C

由表 5 可以得到，以北京为代表的直辖市，因为经济发达，且人口密度较大，

因此人口结构一直不合理，分析近五年的数据，其人口结构都处于最低的等级，山东、浙江、四川、辽宁等地方则可能是因为经济发展不均衡，且人口密度相对较大，农业人口与非农业人口比重失调所造成的人口结构不合理，而一直处于 C 或 D 等级。广东省则是在这些年过程中，逐渐向人口结构合理的方向发展，贵州则与其相反。新疆一直处理人口结构适中的情况，海南作为一个海岛，经济发展较好，人口密度较小，因为人口结构比较合理。

用所建立人口结构发展综合评价指标体系，对题目所给出的附件中国 A 地区进行分析，可以得出 A 地区在 2014 年的评分为 14.57977，所以其人口结构的等级为 C，较为合理。

5.3 对人口结构变化的综合预测模型

在本小问中，要预测人口结构的变化，出生率、死亡率和年龄比例以及性别比例是必须考虑的因素。在前文中的指标中所提到的非农业人口占比体现的是地区经济对人口结构的影响，因此在这里不作考虑。

5.3.1 出生率

题目附件所给出的 A 地区的数据，由于地区政策的影响，因此在 2010 年以后进行了区域划分，使得出生率发生了较大的变化，2010 年以后的出生率变化如下图 2 所示

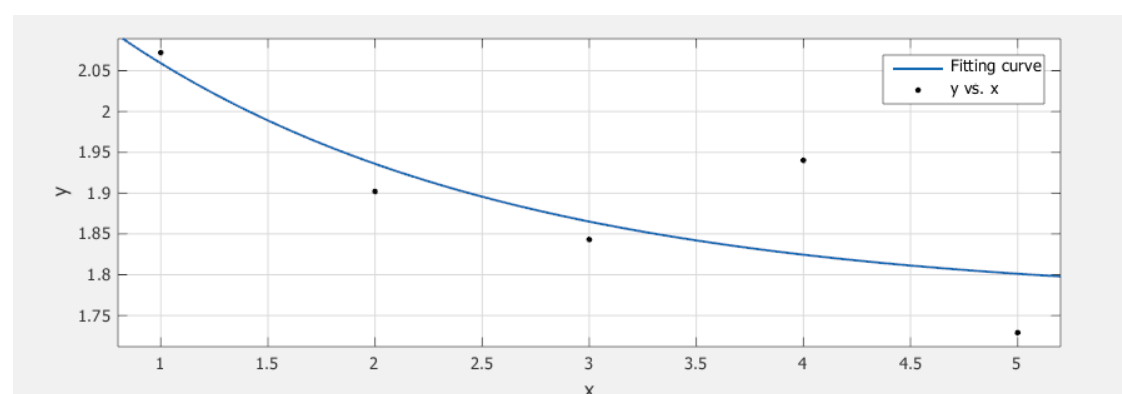


图 2 A 地区 2010 至 2014 年出生率拟合曲线

因此，我们假设此函数曲线为指数函数， $y = a \cdot \exp(-b \cdot x) + c$ ，用 MATLAB 拟合函数曲线结果得出 $a=0.5042$ ， $b=-0.5548$ ， $c=1.77$ 。

因此得出，出生率的函数为：

$$y = 0.5042 \cdot \exp(-0.5548 \cdot x) + 1.77 \quad (4)$$

预测自 2014 年起以后 20 年的出生率如下表 6 所示

表格 6 A 地区 20 年内的出生率预测值

2015 年	2016 年	2017 年
17.07	16.75	16.47
2018 年	2019 年	2020 年
16.23	16.02	15.84
2021 年	2022 年	2023 年
15.55	15.43	15.32
2024 年	2025 年	2026 年
15.23	15.15	15.08
2027 年	2028 年	2029 年
15.02	14.97	14.93
2030 年	2031 年	2032 年
14.89	14.86	14.83
2033 年	2034 年	
14.8	14.78	

出生率的变化趋势如下图 3 所示：

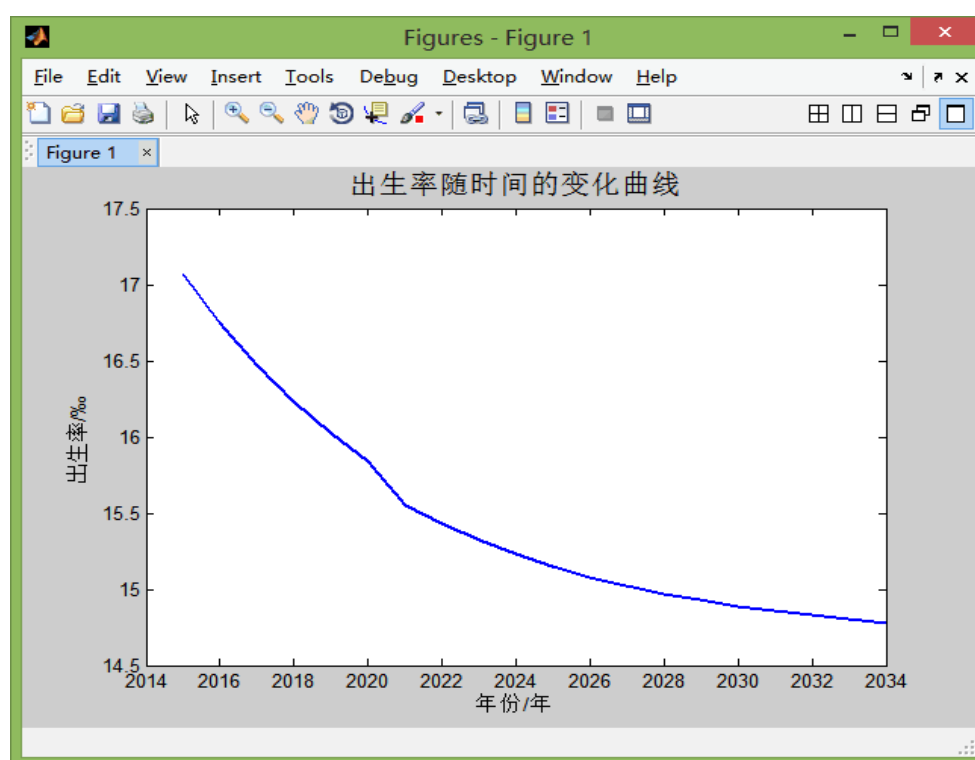


图 3 A 地区未来 20 年出生率变化趋势

因此，分析可以可知，如果在不实行二胎政策时，A 地区未来二十年的出生率逐渐降低，从 2015 年的 17.07 变化到 14.8 左右，且会变化越来越缓慢。

5.3.2 死亡率

对于死亡率的度量一般都是以年龄为自变量的函数，常用的死亡率模型大致

为以下几类，这些模型当中参数 x 表示年龄， $\mu(x)$ 表示的是死亡率。

Compertz 模型：

$$\mu(x)=\alpha \cdot e^{\beta x}.$$

Makeham 模型：

$$\mu(x)=\alpha \cdot e^{\beta x}+\gamma.$$

Weibull 模型：

$$\mu(x)=\alpha \cdot e^{\beta}.$$

Kannisto 模型：

$$\mu(x)=\frac{\alpha \cdot e^{\beta x}}{1+\alpha \cdot\left(e^{\beta x}-1\right)}.$$

通过最小二乘法分别对上述模型的死亡率变化情况进行拟合，根据误差情况在不同的年龄阶段选择合适的模型，表 7 中分别列出了各个模型拟合均方根误差。

表 7 2010 年黑龙江人口死亡率各模型拟合均方根误差比较

年龄段	<i>Compertz</i> 模型	<i>Makeham</i> 模型	<i>Weibull</i> 模型	<i>Kannisto</i> 模型
0-14 岁	0.00345	0.007	0.0025	0.0029
15-64 岁	0.00975	0.0064	0.0082	0.0058
65 岁及以上	0.0322	0.0235	0.0423	0.0232

从上表的数据分析可以看出，在用不同的函数对死亡率进行拟合时，不同模型拟合效果不同，本文中最终决定分段死亡率模型的数学形式为：

$$\mu(x)=\left\{\begin{array}{ll} \alpha \cdot e^{\beta} & , 0 < x \leq 14 \\ \frac{\alpha \cdot e^{\beta x}}{1+\alpha \cdot\left(e^{\beta x}-1\right)} & , x > 15 \end{array}\right. \quad (5)$$

上式中参数的值在 *MATLAB* 中计算出来为：

$$a=2.5152 e-10 ; b=4.4886 ; \alpha=3.3347 e-4 ; \beta=0.0702 .$$

为了确定不同年龄阶段人群未来死亡率可能的变化趋势，可以对死亡率 $\mu(a, t)$ 采用分段加权的方法，即加入对不同年龄阶段设定一个合理的衰减因子，第一部分是婴幼儿时期，在这一阶段的衰减因子的值相对比较大，第二部分是青年时期，本文中设定此时期的人群的死亡率不发生变化，第三部分是老年时期，随着医疗技术的进步，死亡率也会有较大的降低，因此说老年时期的衰减因子相应的也会比较大。

则 $t+1$ 年代 i 岁人口的死亡率可以用下面的公式表示：

$$\mu_i(t+1)=\begin{cases}(1-0.001)\cdot\mu_i(t) & ,0\leq i\leq 5 \\ \mu_i(t) & ,6\leq i\leq 50. \\ (1-0.001)\cdot\mu_i(t) & ,i\geq 51\end{cases} \quad (6)$$

用公式(6)求得 A 地区各个年龄段的死亡率如表 8.

表格 8 A 地区各年龄组人口死亡率预测值 (‰)

年龄组 (岁)	2015-2018	2019-2022	2023-2026	2027-2030	2031-2024
0-14	2.09	1.87	1.55	1.48	1.43
15-64	3.70	3.57	3.31	3.03	2.88
≥65	115.50	109.43	107.78	104.49	101.88

5.3.3 性别结构的灰色预测模型

对 A 地区自 1949 年以来至 2014 年的每一年的男女比例计算, 得到数列的计比, 对原始数据用公式进行级比检验

$$\lambda(k)=\frac{x^{(0)}(k-1)}{x^{(0)}(k)}, k=2,3,\cdots,n. \quad (7)$$

得到附件二的数据, 可知每个级比都落在了可容覆盖区间 $X=(e^{\frac{-2}{n+1}}, e^{\frac{2}{n+1}})$

因此可以对数据建立 GM(1, 1) 进行灰色预测, 所以定义 GM(1, 1) 的灰色微分方程模型为

$$x^{(0)}(k)+az^{(1)}(k)=b, \quad (8)$$

其中, $x^{(0)}(k)$ 称为灰导数, a 为发展系数, $z^{(1)}(k)$ 为白化背景值, b 为灰作用量。

用回归分析求得 a, b 的值分别为 0.0027、1.1946, 因此得到白化模型。

$$\frac{dx^{(1)}(t)}{dt}+0.0027x^{(1)}(t)=1.1946, \quad (9)$$

由方程 (9) 算出自 1949 年至 2014 年的性别比例, 用下列公式计算相对残差和级比偏差

$$\varepsilon(k)=\frac{x^{(0)}(k)-\hat{x}^{(0)}(k)}{x^{(0)}(k)}, k=1,2,\cdots,n, \quad (10)$$

计算级比偏差

$$\rho(k)=1-\frac{1-0.5a}{1+0.5a}\lambda(k), \quad (11)$$

对于所有的 $|\varepsilon(k)|<0.1$, $|\rho(k)|<0.1$, 所以认为此灰色预测模型很好的预测数据。得到自 2014 年起至 2035 年的性别比例预测数据如下表 9 所示

表 9 A 地区性别比例预测数据

2015 年	2016 年	2017 年
1.059498184	1.056641397	1.053792313
2018 年	2019 年	2020 年
1.050950912	1.048117171	1.045291072
2021 年	2022 年	2023 年
1.042472593	1.039661713	1.036858413
2024 年	2025 年	2026 年
1.034062671	1.031274468	1.028493782
2027 年	2028 年	2029 年
1.025720594	1.022954884	1.020196631
2030 年	2031 年	2032 年
1.017445816	1.014702417	1.011966416
2033 年	2034 年	
1.009237792	1.006516525	

由上表可以得出以下曲线图 4:

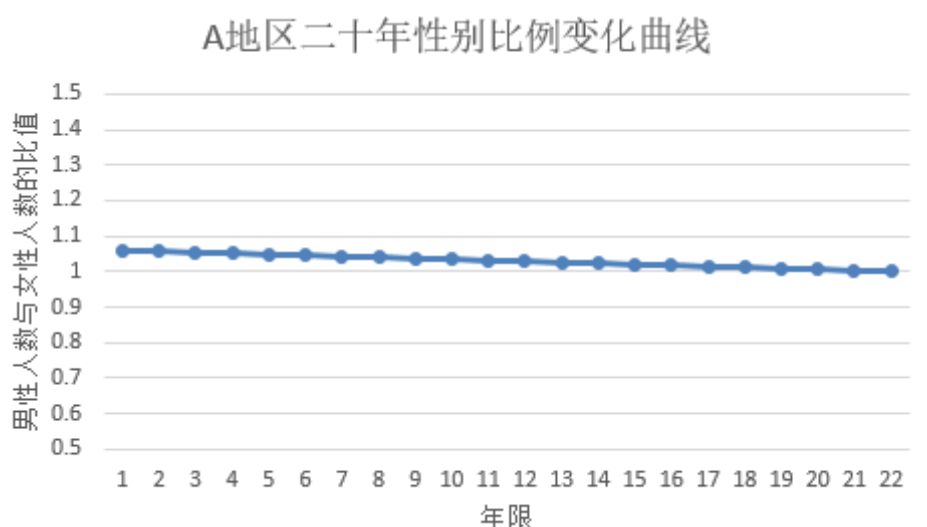


图 4 A 地区性别比例变化曲线

将二十年的性别比例做出曲线，可以更直观的看出，性别比例趋向于 1。

5.3.3 预测年龄结构的房室模型

年龄结构被划分为 0-14 岁，15 岁-64 岁，和 65 及以上三个阶段。我们将这三个阶段的人口数目定义为 x_1 、 x_2 、 x_3 。随着时间的变化，新出生的人数、死亡人数以及从 x_1 中变化到 x_2 的人数都会导致 x_1 的变化，同理 x_2 、 x_3 也会发生变化建立房室模型如下图 5 所示：

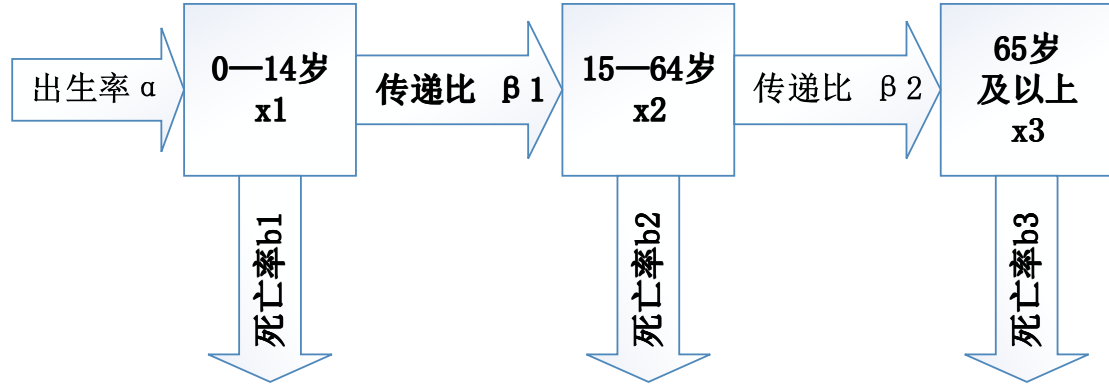


图 5 房室模型

传递比即为从一个年龄组到另一个的传递比例，因此，可以得出微分方程如下：

$$\begin{aligned}\frac{dx_1}{dt} &= x_1(\alpha - b_1 - \beta_1) \\ \frac{dx_2}{dt} &= x_2(-b_2 - \beta_2) + \beta_1 x_1 \\ \frac{dx_3}{dt} &= -x_3 b_3 + \beta_2 x_2\end{aligned}\quad (12)$$

最终解得，

$$\begin{aligned}x_1 &= c_1 \exp((\alpha - b_1 - \beta_1)t) \\ x_2 &= \frac{\beta_1 x_1 - c_2 \exp(-t(b_2 + \beta_2))}{b_2 + \beta_2} \\ x_3 &= \frac{\beta_2 x_2 - c_3 \exp(-tb_3)}{b_3}\end{aligned}\quad (13)$$

5.3.4 问题求解

若 A 地区未实行全面二孩政策，则可以用房室模型最终得出二十年内的人口结构变化。若考虑让 A 地区实行全面二胎政策，则会影响人口的出生率，但由于短时间内医疗水平不会发生大的变化，因此死亡率不会发生大变化。

性别比例预测在 20 年内趋近于 1，且全面开放二胎政策，对性别比例不会构成大的影响。因此人口结构的变化只考虑年龄结构的变化即可。

对于三个年龄段， x_1 、 x_3 的数目不能持续增长，会增加青壮年抚养压力， x_2 的数据不能一直减少，否则会增加使得人口出现负增长。因此 当有一个出生率使得年龄结构在短时间不会发生改变，且使得 $x_1 < x_2, x_3 < x_2$ ，此时的出生率就是我们所求的最优出生率。

求得结果如下表 10 所示：

表 10 两种方案下年龄结构比例 (%)

年份	未实行二孩政策			实行二孩政策		
	0-14 岁	15-64 岁	65 岁及以上	0-14 岁	15-64 岁	65 岁及以上
2015	16.78384	73.06997	10.14619	19.75165	73.39765	6.850692
2016	17.14141	72.13644	10.72215	20.18617	72.47133	7.342496
2017	17.38151	71.37186	11.24664	20.48788	71.70766	7.804463
2018	17.7102	70.43919	11.85061	20.89024	70.78593	8.323829
2019	17.97531	69.47405	12.55064	21.21733	69.83362	8.949043
2020	18.33676	68.44931	13.21392	21.65589	68.82846	9.51565
2021	18.67666	67.42019	13.90315	22.07231	67.82132	10.10637
2022	19.01666	66.3699	14.61343	22.48565	66.79594	10.71841
2023	19.13302	65.50348	15.36351	22.63704	65.94275	11.42021
2024	19.45305	64.44579	16.10116	23.02561	64.91391	12.06048
2025	19.81769	63.55616	16.62614	23.4639	64.04883	12.48727
2026	20.11192	62.4618	17.42628	23.8218	62.98877	13.18943
2027	20.19442	61.85423	17.95136	23.93715	62.38894	13.67392
2028	20.45895	60.73518	18.80587	24.25976	61.30821	14.43203
2029	20.70649	59.59417	19.69935	24.56228	60.20823	15.22949
2030	20.93437	58.43229	20.63334	24.84212	59.08997	16.06791
2031	21.007	57.45011	21.5429	24.94342	58.13911	16.91746
2032	21.1388	56.10403	22.75718	25.11335	56.84637	18.04027
2033	21.24547	54.7402	24.01433	25.2545	55.53778	19.20771
2034	21.26982	53.21273	25.51745	25.30177	54.07244	20.62578

由表 10 数据画出图 6 三个年龄段的发展趋势图。

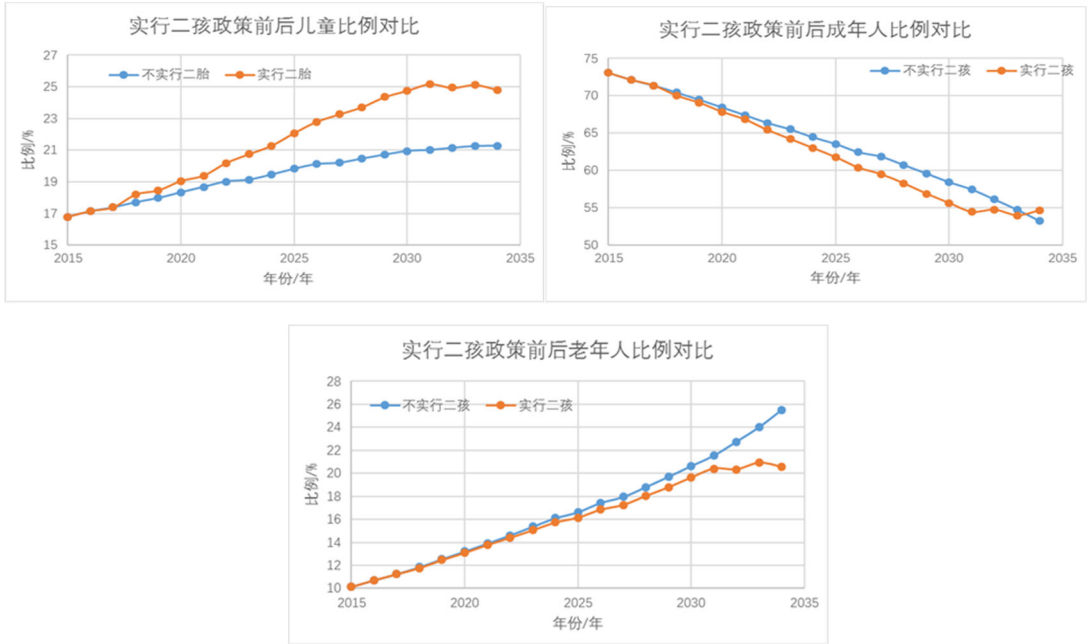


图 6 三个年龄段实行二孩政策前后发展趋势对比

x1, x2, x3 为在不实行二孩政策时, 短时间内一定会依照曲线的发展趋势继

续增加或者减少下去，若没有迁入迁出等外来因素导致其人口结构变化，任其发展下去，最终会使得 A 地区人口结构严重的老龄化。

但是当 A 地区的出生率为 22.35‰ 时，各个年龄段的比例曲线最终会趋向于稳定，且此时 A 地区人口结构老龄化的趋势得到了明显的改善。因此实行二胎政策时的最优出生率为 22.35‰。

5.4 基于生产力水平、文化程度、生育意愿因素影响的改进房室模型

5.4.1 模型建立

实行二孩政策及全面放开生育政策后，由于人们的生育意愿不同，将导致地区人口数量以及人口结构的改变。生育意愿是人们在一定制度、经济、社会环境下对家庭拥有最佳孩子数量和性别的判断。生育意愿往往决定着人们的生育行为，但最终生育行为却取决于社会经济和制度环境的允许范围。虽然如此，生育意愿仍从社会观念和心理角度反映出了人们对自己所处制度、社会、经济环境和自然资源与环境的认识，生育意愿的转变恰恰是这些方面的综合反映

基于对问题的分析，本文中问题三中所建立的人口预测模型进行改进，从而得出生育政策变化后地区人口结构变化趋势预测模型。

生育政策改变后，有前面的分析可以知道由于人们生育意愿，将导致人口数量以及人口结构的改变；而人口结构在变化的过程中又受到人们的生产力水平、生育意愿、文化程度的影响。因此，本文在改进人口预测模型时，主要考虑生产力水平、生育意愿的影响。

通过查阅资料，我们知道人口结构模型的系数与生产力水平成反比，与生育意愿成正比，与文化程度成反比。因此有：

$$\varphi = \frac{1}{a_1} + a_2 + a_3$$

其中， φ 表示人口模型系数； a_1 、 a_2 、 a_3 为常数，且是 0-0.5 之间的小数，分别表示是三个因素的对系数的影响。

在不同政策下，人口出生率也会受到不同的影响，因此，本文中对我们的出生率预测函数进行如下调整：

$$\alpha' = \alpha \cdot \delta_i$$

其中， α' 表示调整后的出生率； δ_i 表示调整系数，二胎政策以及全面放开计划生育政策的调整系数是不相同的。

因此，问题三中的模型可以改进为

$$\begin{aligned}\frac{dx_1}{dt} &= x_1(\alpha' - b_1 - \beta_1 + \varphi) \\ \frac{dx_2}{dt} &= x_2(-b_2 - \beta_2) + \beta_1 x_1 \\ \frac{dx_3}{dt} &= -x_3 b_3 + \beta_2 x_2\end{aligned}\quad (14)$$

5.4.2 问题 4 求解

在二孩政策下，针对 A 地区的当前状况来求解其年龄结构。我们假设生育意愿和文化程度是在 0-1 之间的一个随机的值，生产力水平我们用 2014 年的非农业人口占比来确定。 δ_i 取值为 1，利用公式 (14)，求得 A 地区在实行二孩政策时，10-20 年的年龄结构变化。数据如表 11 所示：

表 11 A 地区未来 10—20 年的年龄结构比例 (%)

年份	0-14 岁	15-64 岁	65 岁及以上
2024	19.27196	67.21819	13.50984
2025	19.50382	66.54412	13.95206
2026	19.39332	65.38187	15.22481
2027	19.2954	64.50537	16.19923
2028	19.23246	63.74893	17.01861
2029	19.56989	63.09158	17.33853
2030	20.11666	62.0913	17.79204
2031	20.22074	61.93157	17.84769
2032	19.97678	60.92208	19.10114
2033	19.27196	67.21819	13.50984
2034	19.50382	66.54412	13.95206

由表中数据可以得出，10 年—20 年中 A 地区的儿童组的人口比例在 19.5% 附近波动，青壮年以及老年组的人口在虽然变化较大，但是都是在某个区间内震荡。若 A 地区以此情形发展下去，对比问题三中的表 11 未实行二孩政策时的年龄结构，可以看出实行二孩政策可以缓解 A 地区人口老龄化问题。

总人口的变化曲线如下：

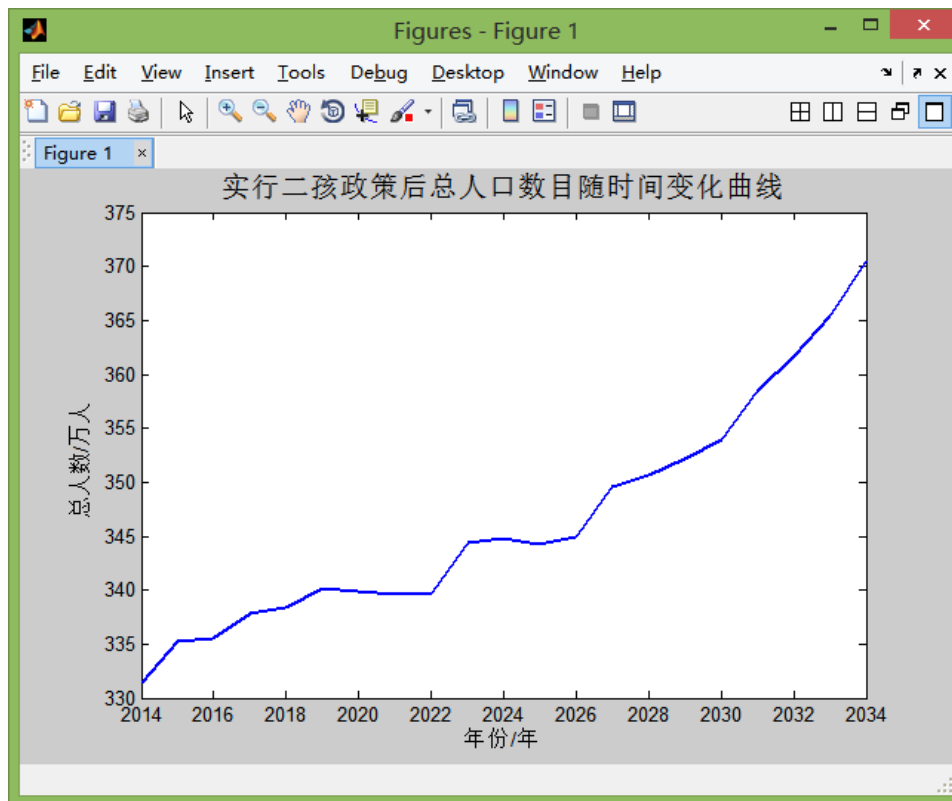


图 7 实行二孩政策后总人口数目随时间的变化曲线

由上图我们可以看出：实行二孩政策后，总人口数目随时间的增加而增加，并且，前 10 年的增长趋势缓慢，后 10 年的人口变化趋势加快，并且在 2034 年总人口数量达到 370.53 万。

5.4.3 问题 5 求解

若国民经济运行稳定，则可以不考虑经济因素对出生率的影响，则上述模型中的 $\varphi = a_2 + a_3$ ，且全面开放生育政策，则使得对一个的出生率的系数变大，使得 δ_i 为一个 (1, 2) 之间的小数，这里我们取 1.5。最终得出表 12 的数据

表 12 全面生育政策下的 A 地区未来 20 年人口结构比例变化 (%)

年份	0-14 岁	15-64 岁	65 岁及以上
2015	17.35827	72.5355	10.10623
2016	17.53597	72.34794	10.11608
2017	17.58948	72.10829	10.30223
2018	18.03648	71.65846	10.30505
2019	18.04382	71.34807	10.60811
2020	18.12076	71.26825	10.61098
2021	18.44738	70.80108	10.75153

2022	18.62976	70.54613	10.8241
2023	18.7095	70.3278	10.9627
2024	19.45961	69.50394	11.03644
2025	19.52634	69.39559	11.07807
2026	19.74452	69.13249	11.123
2027	19.69925	68.59388	11.70687
2028	20.01483	68.28681	11.69837
2029	20.18574	68.05388	11.76038
2030	21.84609	66.37741	11.7765
2031	22.66405	65.50928	11.82667
2032	22.74014	65.37589	11.88397
2033	17.35827	72.5355	10.10623
2034	17.53597	72.34794	10.11608

由上表画出各个年龄段的比例变化趋势如图所示：

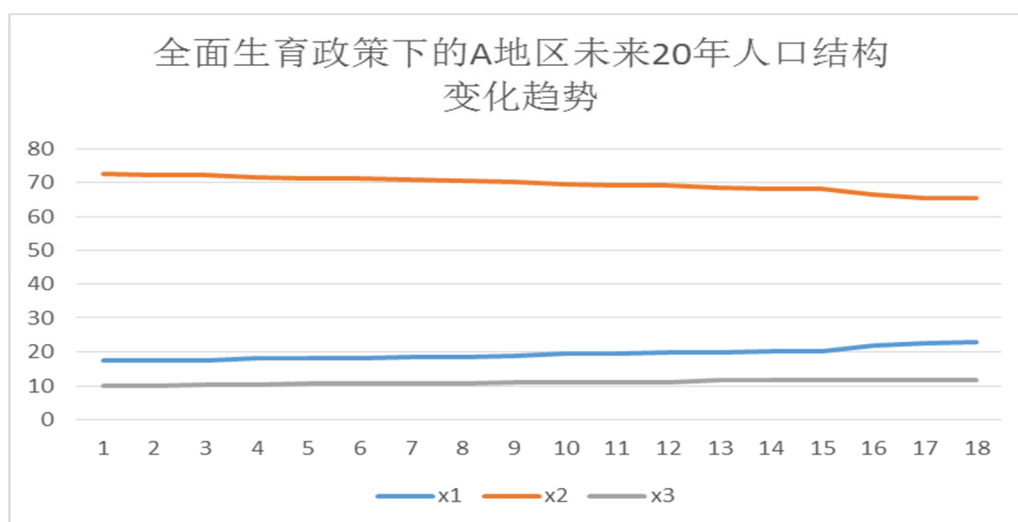


图 8 全面放开生育政策下人口结构变化趋势

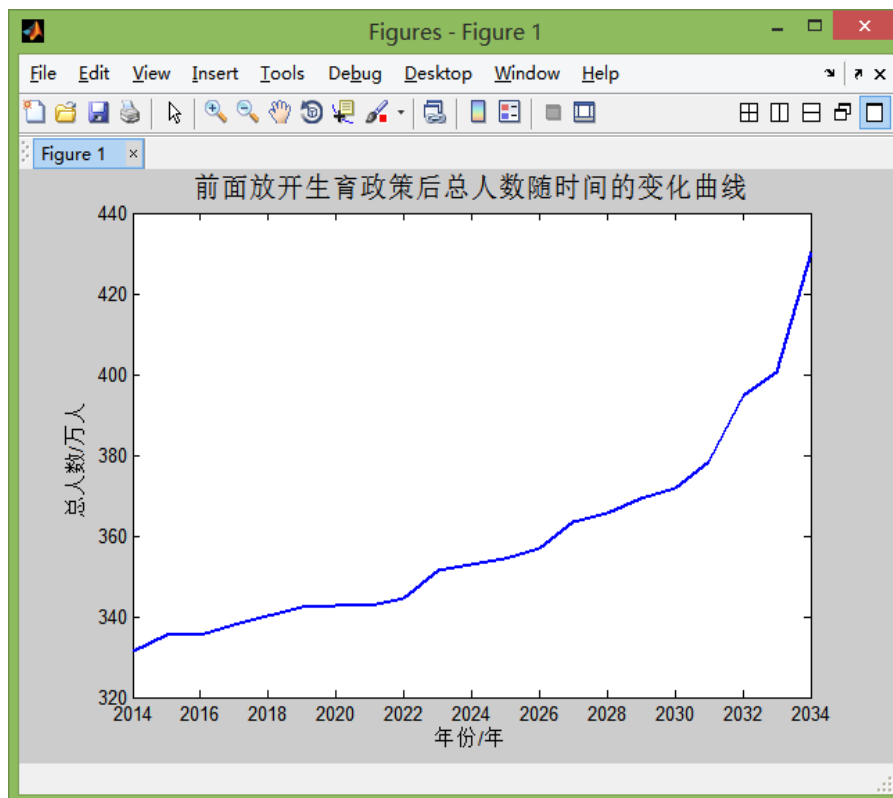


图 9 前面放开生育政策后总人数随时间的变化曲线

由图 8、图 9 可以得到，全面开放生育政策时，在国民经济运行稳定的前提下，儿童组的比例比问题 4 中，开放二胎政策的比例下降的要慢，如果继续预测，应该会趋于稳定状态。说明在经济稳定的前提下，全面开放生育政策，可以更好的达到预期的效果。

六、模型优缺点

6.1 优点

- ✚ 用变异系数法求出各个人口结构的权重，指标取值差异越大的指标更能反映被评价单位的差距，因此更加客观。
- ✚ 本文用灰色预测模型和灰色关联度分析模型，其特点是按发展趋势做分析，因此对样本量的多少没有过多的要求，也不需要典型的分布规律，而且计算量比较小，其结果与定性分析结果会比较吻合。
- ✚ 房室模型可以同时处理多个因变量并且容许自变量和因变量含测量误差，能够同时估计因子结构和因子关系，也能容许更大弹性。
- ✚ 最小二乘法求得出生率和死亡率的拟合曲线，可以简便地求得未知的数据，

并使得这些求得的数据与实际数据之间误差的平方和为最小。

6.2 缺点

- ✚ 本文用变异系数法确定人口结构的权重，具有一定的主观性，当平均值接近于 0 的时候，微小的扰动也会对变异系数产生巨大影响，因此造成精确度不足。
- ✚ GM(1, 1)的预测过程中，用 $\hat{x}(1)$ 作为已知条件来求解预测公式缺乏严格的理论依据，因此解出的不一定是最佳的预测公式。

七、参考文献

- [1]李晖, 陈锡康. 基于人口投入产出模型的中国人口结构预测及分析[J]. 管理评论, 2013. 2502:29-34.
- [2]许锋, 冯健. 城市总体规划人口结构预测分析及其规划启示——以淄博市为例[J]. 城市发展研究, 2006, 06:35-41.
- [3]侯银莉. 湖南省人口数量及结构的预测研究[D]. 中南大学, 2010.
- [4]蒋赛. 黑龙江省人口数量与结构的预测研究[D]. 哈尔滨工业大学, 2013.
- [5]宣丹萍. 浙江省人口年龄结构预测研究[D]. 杭州电子科技大学, 2013.
- [6]解保华, 陈光辉, 孙嘉琳. 基于 Leslie 矩阵模型的中国人口总量与年龄结构预测[J]. 广东商学院学报, 2010. 25;No. 11003:15-21.
- [7]虞丽萍. 人口年龄结构模型建模和预测[D]. 上海交通大学, 2007.
- [8]王学萌, 郝永红, 黄登宇. 中国人口结构的灰色动态预测[A]. 中国优选法统筹法与经济数学研究会. 2004:4.
- [9]王学萌, 郝永红, 黄登宇. 中国人口结构的灰色动态预测[A]. 中国优选法统筹法与经济数学研究会. 2004 年中国管理科学学术会议论文集[C]. 中国优选法统筹法与经济数学研究会:, 2004:4.
- [10]翟振武, 张现苓, 靳永爱. 立即全面放开二胎政策的人口学后果分析[J]. 人口研究, 2014, v. 38;No. 20602:3-17.
- [11]张迎春, 侯园园, 韩晓庆. “单独”二胎政策条件下人口年龄结构预测研究[J]. 经济统计学(季刊), 2014, No. 302:135-146.
- [12]安然. 北京市二胎政策对人口老龄化的影响[D]. 中国地质大学(北京), 2015.
- [13]韩晓庆. 基于 Leslie 模型中国未来人口策略模拟研究[D]. 东北财经大学, 2012.

附录

附录 1:

指标	出现次数
年龄结构	67
性别比例	65
出生率	74
死亡率	75
非农人口占比	56
人口密度	51
人口总数	45
人口迁移量	43
就业人口承载能力	23
经济自由度	13
社会保障	24
收入差距	23
居住条件	15
基础设施	23
环境污染	10

附录 2:

十个省份的数据——来自国家统计局（以 excel 文件附加）

[illegible]

附录 3:

误差分析

X0k	残差	相对误差(小数)	极比偏差
1.133999	0	0	
1.189931	-0.10221	0.102208488	-0.04757
1.186723	-0.11527	0.115269151	-0.01185
1.183523	-0.08786	0.087859265	0.024577
1.180332	-0.01918	0.01918399	0.063129
1.177149	-0.03349	0.033494106	-0.01404
1.173975	0.009138	0.009138276	0.041251
1.170809	0.037932	0.03793204	0.029059
1.167653	0.031578	0.031577772	-0.0066
1.164504	0.028162	0.02816248	-0.00353
1.161364	0.026596	0.026595641	-0.00161
1.158233	0.034898	0.034897585	0.008529
1.15511	0.042117	0.042117089	0.007481
1.151995	0.04262	0.042619804	0.000525
1.148889	0.042635	0.042635065	1.59E-05
1.145791	0.043716	0.043715632	0.001129
1.142702	0.037621	0.037621289	-0.00637
1.139621	0.037338	0.037337638	-0.00029
1.136548	0.037673	0.037672741	0.000348
1.133483	0.035452	0.035451603	-0.00231
1.130427	0.033476	0.033475835	-0.00205
1.127379	0.042704	0.042703633	0.009547
1.124339	0.031978	0.031977904	-0.0112
1.121307	-0.02198	0.021975799	-0.05574
1.118284	-0.02453	0.024527701	-0.0025
1.115269	-0.02116	0.021157503	0.00329
1.112262	-0.02003	0.020025236	0.001109
1.109263	-0.02076	0.02076274	-0.00072
1.106272	-0.02476	0.024755319	-0.00391
1.103289	-0.02446	0.024463375	0.000285
1.100314	-0.02197	0.021971743	0.002432
1.097347	-0.01952	0.019516344	0.002403
1.094388	-0.01363	0.013634922	0.005769
1.091437	-0.0147	0.014699478	-0.00105
1.088494	-0.012	0.012004268	0.002656
1.085559	-0.00931	0.009312965	0.002659
1.082632	-0.00867	0.008665758	0.000641
1.079713	-0.00772	0.00771716	0.00094
1.076802	-0.00568	0.005678565	0.002023
1.073898	-0.01122	0.011222092	-0.00551

1.071003	-0.00842	0.008424966	0.002766
1.068115	-0.00549	0.005489968	0.00291
1.065235	-0.0046	0.004603572	0.000882
1.062363	-0.00456	0.004558491	4.49E-05

附录 4:

```

d11 = [2.09 2.09 2.09 2.09 1.87 1.87 1.87 1.87 1.55 1.55 1.55 1.55 1.48 1.48
1.48 1.48 1.43 1.43 1.43 1.43] ;
d12 = [3.70 3.70 3.70 3.70 3.57 3.57 3.57 3.57 3.31 3.31 3.31 3.31 3.03 3.03
3.03 3.03 2.88 2.88 2.88 2.88];
d13 = [115.50 115.50 115.50 115.50 109.43 109.43 109.43 109.43 107.78 107.78
107.78 107.78 104.49 104.49 104.49 104.49 101.88 101.88 101.88 101.88];
c1 = 55;
c2 = -4.2;
c3 = -3.9;
p1 = 0.01;
p2 = 0.015;
t = [1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20];
birth = [0.01707
0.01675
0.01647
0.01623
0.01602
0.01584
0.01555
0.01543
0.01532
0.01523
0.01515
0.01508
0.01502
0.01497
0.01493
0.01489
0.01486
0.01483
0.0148
0.01478
];
for i=1:1:20
x1 = c1 * exp((birth(i) - d11(i)/1000 - p1) * t(i));
x2 = (p1 * x1 - c2 * exp(t(i)*(d12(i)/1000+p2))) / (d12(i)/1000+p2)
x3 = (p2 * x2 - c3 * exp(t(i)*d13(i)/1000)) / (d13(i)/1000)
end;

```