Modeling and Simulation 建模与仿真, 2022, 11(5), 1309-1319
Published Online September 2022 in Hans. http://www.hanspub.org/journal/mos
https://doi.org/10.12677/mos.2022.115123

基于全国人口普查问题模型的研究与思考

化紫康

周口师范学院，河南 周口

收稿日期：2022年8月16日；录用日期：2022年9月13日；发布日期：2022年9月22日

摘 要

研究全国人口普查问题模型，旨在为政府统计部门核实人口数量和统计人口结构提供重要参考，以做出科学决策。研究对象是人口普查标准时点上的人口总体数量。利用这些数据推理、计算出最大人口数和人口增长率，得到人口普查问题模型。通过数理论推理数理逻辑，用MATLAB进行模型实验，实际数据与预测数据拟合，进行误差微调。经过实验证验并可以对人口数量预测，进行人口结构预测，为政府决策提供数据支撑。

关键词

人口普查，指数模型，人口数量，人口结构

Research and Thinking on the Problem Model of National Census

Zikang Hua

Zhoukou Teachers College, Zhoukou Henan

Received: Aug. 16th, 2022; accepted: Sep. 13th, 2022; published: Sep. 22nd, 2022

Abstract

The purpose of studying the national population census model is to provide important references for the government statistics department to verify the population number and statistical population structure so as to make scientific decisions. The research object is the total population at the standard time point of the census. By reasoning these data, the maximum population and the population growth rate are calculated, and the population census model is obtained. The model experiment is carried out by using MATLAB through mathematical theory and mathematical logic. The actual data is fitted with the predicted data, and the error is fine tuned. After experimental verification, it can predict the population number and the population structure, and provide data support for government decision-making.

文章引用: 化紫康. 基于全国人口普查问题模型的研究与思考[J]. 建模与仿真, 2022, 11(5): 1309-1319.
DOI: 10.12677/mos.2022.115123

Keywords

Population Census, Exponential Model, Population Size, Population Structure

Copyright © 2022 by author(s) and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

1. 引言

第七次全国人口普查的主要数据结果显示 2020 年中国人口共 14.12 亿人，与 2010 年的 13.40 亿人相比，有所增加。2020 年全国人口中，男性人口为 723,339,956 人，占 51.24%；女性人口为 688,438,768 人，占 48.76%，人口结构也在变化。第七次全国人口普查的发布使得全国人口普查话题再次引起人们的关注。全国人口普查是由国家组织依法对全国现有人口普遍地、逐户逐人地进行一次全项调查登记，普查重点是掌握各地现有人口的变化、性别比例及城乡人口等数据，以便国家制定下一步发展政策。第七次全国人口普查是我国人口发展进入关键期开展的一次重大国情国力调查。由于我国人口总量很大，每次进行的全国人口普查工作都要花费大量的人力、物力和财力，如果能用数学建模的方式给出人口普查各种数据的有效预测将是一件非常有意义的事情。

齐婧[1]指出在人口普查过程中的清查摸底工作是对普查对象进行精确定位、落实人口普查工作质量的重要基础，许珂等[2]在研究新时代中国人口结构的发展态势研究中指出我国人口规模面临负增长趋势、人口老龄化程序加剧将造成人口红利衰减、地域人口分布不均衡等，这都需要用人口普查模型进行人口结构的统计划分。李白咏[3]在《美国人口普查积弊已久，AI 有望抹平数字鸿沟》一文中指出用 AI 来挽救美国人口普查败局。陈友华[4]进行了人口普查数据与行政统计数据偏离现象研究，谭永宏[5]等进行了基于多项式神经网络模型的人口观测方法研究，胡桂华[6]等采取数理模型方法研究组合式三系统估计量及其与之相关的抽样及估计问题。这些研究都指出要采用数字化进行人口普查，以提供科学、精准的决策数据。

综上所述，根据全国前七次人口普查的主要数据结果，通过数理逻辑建立人口普查预测模型，进行人口数量预测，并能进行人口结构的精确划分，推进人口普查的数字化进程。

2. 相关技术

2.1. 问题描述

我国目前的全国人口普查间隔约是 10 年一次。新中国成立(1949 年)至今，一共进行了七次全国人口普查，这七次人口普查时间如下表 1 所示。

Table 1. Time points for the seven censuses
表 1. 七次人口普查的时间点

第一次	第二次	第三次	第四次	第五次	第六次	第七次
1953.6.30 24 时	1964.6.30 24 时	1982.7.1 0 时	1990.7.1 0 时	2000.11.1 0 时	2010.11.1 0 时	2020.11.1 0 时

文献[1]给出的 7 次人口普查数据作为目标样本数据进行建模与预测分析，中国历次人口普查数据如

表 2 所示。

Table 2. China's previous census population data table
表 2. 中国历次人口普查人口数据表

年份	1953	1964	1982	1990	2000	2010	2020
人口总量	5.94 亿	6.95 亿	10.08 亿	11.34 亿	12.66 亿	13.40 亿	14.12 亿
增长率	23‰	27.8‰	15.68‰	14.39‰	7.58‰	4.8‰	1.4‰

人口普查数据的散点图如图 1 所示。

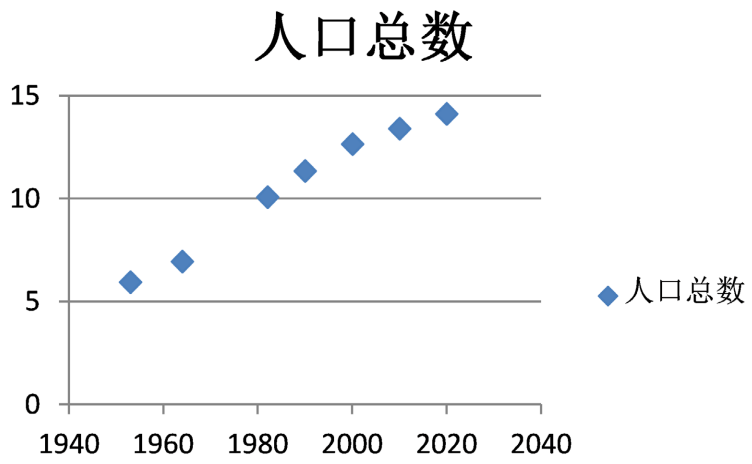


Figure 1. Scatter plot of seven census data in China from 1949 to 2021
图 1. 我国在 1949 年到 2021 年进行的七次人口普查数据散点图

从图 1 中可以看出，我国从 1949 年到 2021 年人口总数一直处于递增趋势，1964 到 1982 的增幅最大，2000 到 2010 的增幅放缓。说明人口的增长会受人类生活所需要的环境和资源等影响，增长率并不会是一直增长，会随人口增长而呈现减小趋势。

2.2. 模型描述

人口预测模型最常见的是指数增长模型(Malthus 模型)与阻滞增长模型(Logistic 模型) [7]。Malthus 模型简单的认为人口增长率 r 不变，Logistic 模型则将人口增长率 r 记成与人口总数 x 有关的函数，并且随着 x 的增大减小，是符合自然规律的，因此在此选用的是 Logistic 模型。

1) Logistic 模型

假设 $r(x)$ 是 x 的线性函数，即

$$r(x)=r-sx(r>0,s>0) \tag{1}$$

这里的 r 表示人口极少时(理论上设 $x=0$)的增长率，即人口不受环境和资源限制的固有增长率。s 为待求系数，为了明确参数 s 的意义，引入最大人口容量 x_m，即当 $x=x_m$ 时，人口增长率为 0。

从图 1 可以看出，1964 年之后 r 是 x 的递减函数，以年份作为时间 t，则以 1964 年作为 $t=0$ 的时刻，记作 $x(0)=x_0$。根据人口增长率为常数 r，则可知单位时间内 $x(t)$ 的增量等于 r 乘以 $x(t)$，考虑 t 到 $t+\Delta t$ 时间内人口的增量，则有

$$x(t+\Delta t)-x(t)=r*x(t)*\Delta t \tag{2}$$

令 $\Delta t \to 0$，得到微分方程：

$$\frac{\mathrm{d}x}{\mathrm{d}t} = rx = r(x)x \tag{3}$$

由于当 $x = x_m$ 时，人口增长率为 0，可得

$$r(x_m) = 0 \Rightarrow s = r/x_m \Rightarrow r(x) = r\left(1 - \frac{x}{x_m}\right)$$

于是(1)式可改写为

$$r(x) = r(1 - x/x_m) \tag{4}$$

把(4)式带入(3)式可得

$$\begin{cases} \dfrac{\mathrm{d}x}{\mathrm{d}t} = r\left(1 - \dfrac{x}{x_m}\right)x \\ x(0) = x_0 \end{cases} \tag{5}$$

2) 人口预测模型

假设：一些大型自然灾害不考虑在内，如战争，地震等。中国实行的生育模式一直不变。医疗水平的变化对人口数量没有影响。

在上述假设下，认为人口增长有最大值 x_m 和人口的固有增长率 r_0。当人口增长到 x_m 附近，人口将保持这个水准不会有大的变动，数学公式：

$$\begin{cases} \dfrac{\mathrm{d}x}{\mathrm{d}t} = r_0\left(1 - \dfrac{x}{x_m}\right)x \\ x(0) = x_0 \end{cases} \tag{6}$$

在(6)式中，x 是人口数量，x_0 是初始人口数量，

由变量分离法(由 $y' = \dfrac{\mathrm{d}y}{\mathrm{d}x}$ 可以把 x，y 的微分和自变量相互分离。然后将其化为 $g(y)\mathrm{d}y = f(x)\mathrm{d}x$ 的形式。变量分离到等式两端时，两边同时积分，得出通解 $g(y) = f(x) + c$)可以求得方程(6)的通解：

$$\frac{x}{x_m - x} = c\mathrm{e}^n \tag{7}$$

利用初始条件得：

$$c = \frac{x_0}{x_m - x_0} \tag{8}$$

把 c 代入通解并化简可得

$$x(t) = \frac{x_m}{1 + \left(\dfrac{x_m}{x_0} - 1\right)\mathrm{e}^{-n}} \tag{9}$$

(9)式可以简写为：

$$x(t) = \frac{x_m}{1 + a\mathrm{e}^{-bt}} \tag{10}$$

其中 $a=\frac{x_m}{x_0}-1$， $b=r$ 。

3) 算法步骤

Step 1：计算每一年的增长率。

满足人口增长的微分方程为 $\mathrm{d}x/\mathrm{d}t=r*x$ 以及初始条件 $x(0)=x_0$ 。先对人口数据进行数值微分求出 x 在各点导数的近似值(根据数值微分中点公式)为：

$$x'(t_k)=(x_{k+1}-x_{k-1})/2\Delta t\ (k=1,2,\cdots,n-1) \tag{11}$$

其中 k 为年序号，从 1 开始起；t_k 为年份号；现约定 Δt 为 1， $x'(t_k)$ 对应 1964~2010 这 5 年的增长率。另外第一次人口普查(1953 年)的增长率为：

$$x'(t_0)=(-3x_0+4x_1-x_2)/2\Delta t \tag{12}$$

第七次人口普查(2020 年)的增长率为：

$$x'(t_n)=(x_{n-2}-4x_{n-1}+3x_n)/2\Delta t \tag{13}$$

Step 2：使用非线性拟合法来计算参数 r_0 以及 x_m 。

采用最小二乘法，设拟合直线的公式为 $y=k*x+b$ ，其中拟合直线的斜率为 $k=\frac{\sum xy-n\overline{xy}}{\sum x^2-n\overline{x^2}}$ ，计算出斜率之后，根据($\bar{x}$ ， $\bar{y}$)和已知斜率可以利用待定系数法求出截距 b，即为固有增长率 r_0 。

由于人口容量的限制，当人口数量增加时，增长率 r 减小，用 $r(x)=a+b*x$ 表示。即

①当 $x=0$ 时， $r(0)=r$ ，则 $a=r$ 。

②当 $x=x_m$ 时，此时人口不再增长，增长率 $r=0$ ，于是： $b=-r/x_0$ 。

Step 3：根据拟合情况，实际数据与预测数据进行比较，计算出误差值，对预测所得数据进行微调。

3. 数据实验与分析

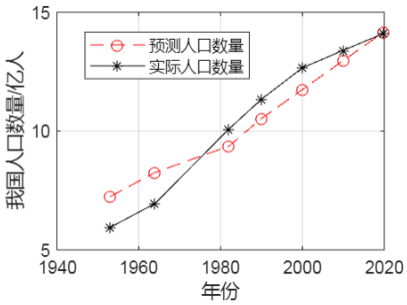
3.1. 全国人口预测

利用 Logistic 模型，可以利用 MATLAB 编写下述程序代码。

```
x=[5.94 6.95 10.08 11.34 12.66 13.40 14.12];%我国人口总数
year1=[1953 1964 1982 1990 2000 2010 2020];
n=length(x);
t=0:1:n-1;
rk=zeros(1,n);
rk(1)=(-3*x(1)+4*x(2)-x(3))/2;
rk(n)=(x(n-2)-4*x(n-1)+3*x(n))/2;
for i=2:n-1
    rk(i)=(x(i+1)-x(i-1))/2;
end
rk=rk./x;
p=polyfit(x,rk,1);
b=p(2); %常数项
a=p(1); %指拟合得到的多项式的一次项系数
```

```
r0=b;
xm=-r0/a;

%输出
pnum=zeros(n,1);
for i=0:1:n-1
    pnum(i+1)=xm/(1+(xm/x(1)-1)*exp(-r0*i));
end

plot(year1,pnum+1.3,'r--o',year1,x,'k-*');

p5=polyfit(year1,x,5);
people_test=polyval(p5,year1)

xlabel('年份')
ylabel('我国人口数量/亿人')
grid on
legend('预测人口数量','实际人口数量')
```

输出结果：

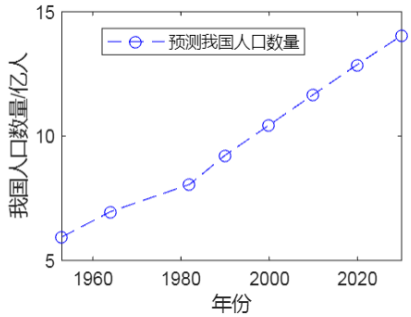


people =

5.9391　　6.9539　　10.0565　　11.3808　　12.6291　　13.4128　　14.1178

通过数据比较和拟合情况，进行微调，预测数据从第三个数开始大致相差 1.3，到第七次人口普查数据完全吻合。预测 2030 年全国人口数量，利用上述代码可以求出。

实验输出结果为。



people =

14.0457

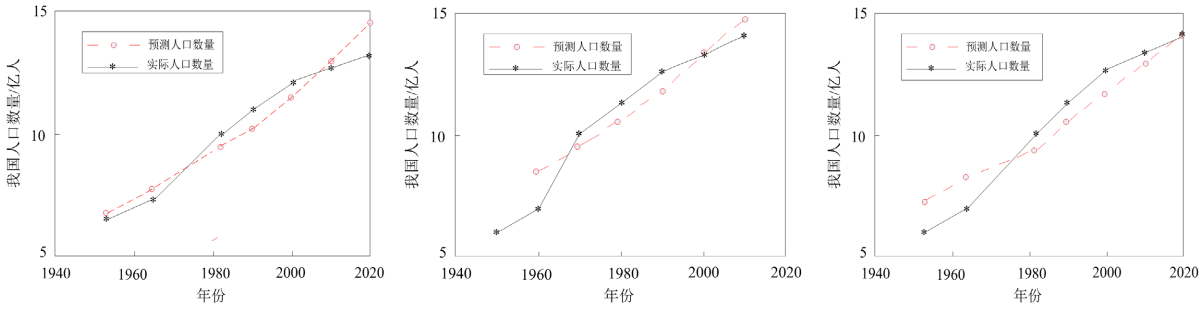
到 2030 年时预测我国人口普查数据约有 14.0457 亿人，若再加 1.3 亿人的误差，大致为 15.3457 亿人。

对表 2 中数据集分别采用马尔萨斯模型、灰色预测模型和回归预测模型进行人口数据的预测。为了方便表示，将这三种方法进行简化表达，如表 3 所列。

Table 3. Comparative experimental model
表 3. 对比实验模型

简化表示	模型
Pfm1	马尔萨斯模型
Pfm 2	灰色预测模型
Pfm 3	回归预测模型

对表 2 中的数据集通过不同的方法进行人口预测得到数据分布如图 2 所示。



(a) Pfm 1 (b) Pfm 2 (c) Pfm 3

Figure 2. Distribution of predicted population and actual population
图 2. 预测人口数量与实际人口数量分布图

从图 2 中可以看出，由 Pfm 1 可以看出预测人口会随着时间的增加而不断增加，不会受到环境和资源等因素的限制，有一定的局限性。由 Pfm2 可以看出灰色模型如果不考虑内在机理，就会出现比较大的错误，由于它是基于模型中参数的模糊性进行研究，所以需要更加专业的处理才能得以运用。总的来说，与其他模型相比，Pfm3 的数据分布与原始数据关联度最高，且能够对指标进行综合，该模型的优势更加明显。

思考：1) 此人口预测模型没考虑自然灾害等未知影响因素，如近期发生的新冠疫情对人口数量的影响。另外人口数据的预测会受到经济增长、国家政策、自然因素和社会因素等综合影响，这些因素在模型中又如何体现。

2) 约十年一次的人口普查是一项系统工程，由于人口数量大、居住地分散、工作转换频率高等客观原因，对预测的 15.3457 亿人进行普查工作的难度是相当困难的，是否需要考虑电子标签等智能数据收集手段。

3) logistic 回归模型的优点是实现简单，容易使用和解释，而且便利观测样本概率分数，广泛的应用于工程问题上。但人口数据的预测具有时代性，因此还要结合贝叶斯分层模型进行先验概率的假设，借助灰色预测模型来验证其因不确定性带来的偏差和局限性，进而得到科学、准确的数据。

4) 人口普查的目的是全面掌握全国人口的基本情况，尤其不同领域需要根据不同的人口结构数据，研究制定人口政策和经济社会发展规划、决策等，人口结构的精准定位是核心问题。

3.2. 河南省人口预测

河南省在 1949 年到 2021 年进行的七次人口普查数据如表 4 所示。

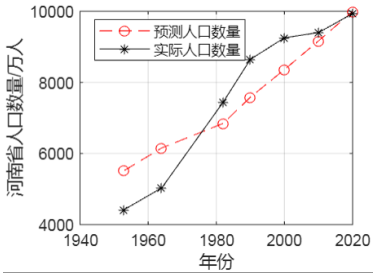
Table 4. Data from seven censuses conducted in Henan Province from 1949 to 2021
表 4. 河南省在 1949 年到 2021 年进行的七次人口普查数据

年份	1953	1964	1982	1990	2000	2010	2020
总人口数量	4420 万人	5032.5 万人	7442.2 万人	8649 万人	9255.8 万人	9402.3 万人	9936.6 万人

代码如下：

```
x=[4420 5032.5 7442.2 8649 9255.8 9402.3 9936.6];%河南省人口总数
year1=[1953 1964 1982 1990 2000 2010 2020];
……
plot(year1,pnum+1100,'r--o',year1,x,'k-*');
……
```

输出结果为：



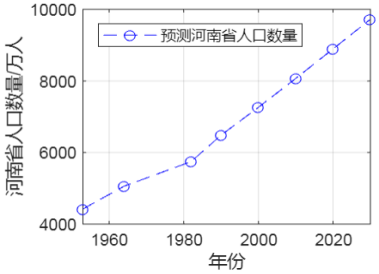

```
people =
    1.0e+03 *
    4.4200    5.0560    5.7457    6.4835    7.2616    8.0694    8.8950
```

通过数据比较和拟合情况，对预测人口数据进行微调，大致相差 1100 万人，到第七次人口普查数据完全吻合。

预测常住人口数量：

```
x=[4420 5032.5 7442.2 8649 9255.8 9402.3 9936.6];%河南省人口总数
year1=[1953 1964 1982 1990 2000 2010 2020 2030];
……
plot(year1,pnum,'b--o');
people=pnum(8)
……
```

输出结果为。



people =

9.7252e+03

到 2030 年时预测河南省人口普查数据约有 9725.2+1100=10825.2 万人，大致 1.082 亿人。依据此模型分别按性别比、城乡比、老龄化比等预测，结果如表 5 所示。

Table 5. Predicted population structure of Henan Province
表 5. 预测河南省人口结构表

年份	1953 年	1964 年	1982 年	1990 年	2000 年	2010 年	2020 年	2030 年
总人口数量	44,214,594	50,325,511	74,422,573	85,509,535	91,236,854	94,029,939	99,365,519	108,250,213
男性	22,745,054	25,490,705	37,949,774	43,900,595	47,075,743	47,493,063	49,831,362	55,082,131
女性	21,469,540	24,834,806	36,472,799	41,608,940	44,161,111	46,536,876	49,534,157	53,168,082
性别比	105.90%	102.60%	104.05%	105.5%	106.60%	106.75%	100.6%	103.6%
城镇人口	3,536,093	5,517,360	10,771,790	11,488,123	17,180,966	26,294,786	35,852,883	38,008,106
农村人口	40,678,501	44,808,151	63,650,783	74,021,412	74,055,887	67,735,153	63,512,635	70,242,107
城乡比	8.69%	12.31%	14.42%	15.52%	23.20%	38.82%	56.45%	54.11%
少年人口率	35.71%	39.76%	34.90%	26.72%	27.07%	30.20%	22.25%	22.63%
老化率	5.13%	4.26%	5.23%	6.84%	6.96%	12.60%	13.99%	14.56%
老化指数	14.37%	10.70%	14.98%	25.60%	25.7%	41.7%	62.8%	70.2%

由表 5 中的数据生成如图 3 和图 4 所示的曲线。

看图 3 与图 4 思考：1) 人口增速放慢。由上述数据可知河南人口从 1953 年到 2030 年，人口数量从 4421 万人一直增长到 1.08 亿人，发展经历了加速、超速、减速、均速等阶段。即与全国人口增长率(23‰、27.8‰、15.68‰、14.39‰、7.58‰、4.8‰、1.4‰)一致，低生育率是当今世界很多国家面临的共性问题之一，作为人口大省的河南如何面对此问题。

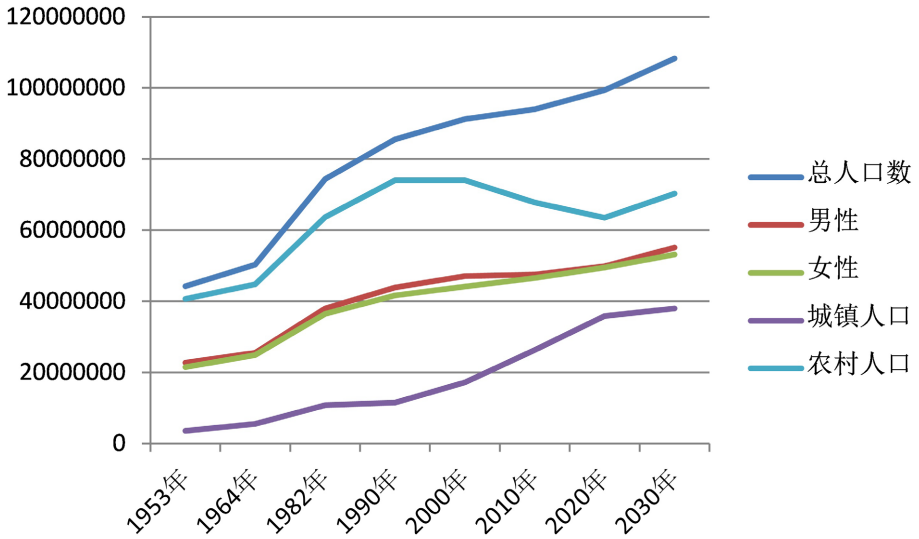


Figure 3. Quantity curve of population structure
图 3. 人口结构数量曲线图

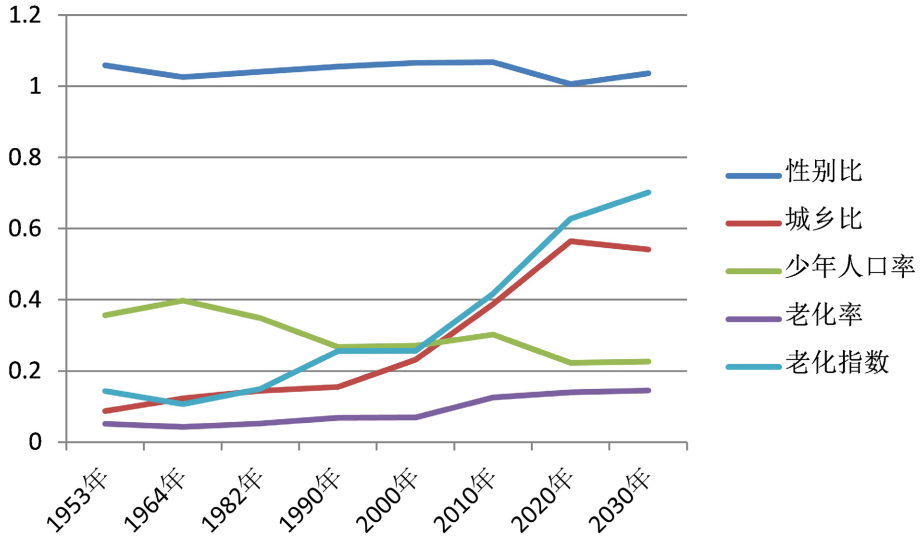


Figure 4. Demographic ratio curve
图 4. 人口结构比率曲线图

2) 老龄化加快。河南人口老龄化速度比较快，少年人口率直线下降，老化率直线上升，导致老化指数越来越大。从表中可以看出 1990 年的老化指数为 25.60%，2010 年的老化指数就高达 41.7%，这标志着河南人口的初期老化阶段，预测 2030 年老化指数为 70.2%，从 2010 年到 2030 年人口老龄化已经进入加速阶段。如此庞大的老龄化队伍，又如何进行资源的调配和协调。

3) 城镇化加速。城乡比例逐年增长，2020 年高达 56.45%，预计 2030 年比率为 54.11%，这样农村人口所占比率越来越小，需要考虑如何协调流动人口，以及城乡资源如何调配。农村人口所占比极小，而且由于农村老龄化现象加剧，会呈现劳动人口从农村向城镇迁移的增速可能放慢，作为农业大省的河南，大量农村的土地资源又该如何管理和利用。

4) 人口流动加快。随着改革开放的深入发展，河南的人口迁移和流动也随之增加，流动人口增加更快。人口流动的原因主要是务工经商和中转，由于河南经济落后，不能形成强大的经济吸引力，迁入人口较少，前来务工经商的人口比重不高， 省际人口迁移模式在近期和中期内不会有大的变化。

4. 结论

通过对人口数量预测模型的研究，并对全国人口数量预测、拟合计算，得到 2030 年全国人口数量，思考人口普查存在的困难，然后又对河南省的人口结构预测并分析，思考由人口预测数据产生的不同人口结构带来的问题。

本文详细地阐述了人口预测模型的推理和计算，实验证明人口预测数量的科学性，并通过人口数量呈现的不同人口结构进行分析、思考。主要结论有以下几点：

1) 此模型是依据指数法、最小二乘法、积分法等数理逻辑进行参数的计算，前后逻辑性很强。

2) 此模型逻辑性强，科学、合理，可加快人口普查的数字化进程，降低人口普查的难度。

3) 借助此模型可以快速定位人口结构，供政府针对不同时期的人口结构做出科学决策。

此模型的推理过程逻辑性强，有一定的实用价值，此人口预测模型没考虑自然灾害等未知影响因素，下一步的工作是对参数进行改进、优化，研究不可避免的未知因素影响下人口预测结果，以供政府统计部门借助人口普查问题模型进行人口结构预测，来完善国家和省人口发展战略和政策体系，促进人口长期均衡发展的迫切需要，科学制定国民经济和社会发展规划，推动经济高质量发展。

参考文献

[1] 齐婧. 对“第七次人口普查”清查摸底工作的几点思考[J]. 上海商业, 2022(6): 235-237.

[2] 许珂, 贺晓迎. 新时代中国人口结构的发展态势研究[J]. 成都行政学院, 2022(2): 33-46.

[3] 李白咏. 美国人口普查积弊已久, AI 有望抹平数字鸿沟[J]. 中国电信业, 2022(5): 67-69.

[4] 陈友华, 孙永健. 人口普查数据与行政统计数据偏离现象研究[J]. 人口与经济, 2022(3): 86-96.

[5] 谭永宏, 曾喆昭. 基于多项式神经网络模型的人口预测方法研究[J]. 数学的实践与认识, 2016, 46(18): 152-158.

[6] 胡桂华, 文婷, 刘誉环. 基于组合式三系统估计量的人口普查净误差估计[J]. 统计与信息论坛, 2022, 37(8): 15-27.

[7] 盛亦男, 顾大男. 概率人口观测方法及其应用——《世界人口展望》概率人口预测方法简介[J]. 人口学刊, 2020, 45(5): 31-46.