

基于线性回归模型、马尔萨斯人口增长模型及 logistic 模型的全国人口预测

赵天伟, 陈惠达* (广东医科大学公共卫生学院, 广东东莞 523808)

摘要: **目的** 为了更好地认识我国人口的数量变化规律, 并对我国未来人口进行预测。**方法** 根据国家统计局官网公布的 1949–2019 年人口数据, 计算人口增长率, 分析人口增长规律; 利用 1949–2014 年的人口数据依次建立线性回归模型、马尔萨斯人口模型及 logistic 模型, 利用相对误差对 3 种模型进行评价, 预测我国 2020–2040 年人口数, 分析未来人口发展变动趋势。**结果** 我国人口增长率呈下降趋势, 1960、1961 年出现人口负增长, 其余年份均为正增长; 通过模型拟合, 对比发现: 线性回归模型存在较大误差, 马尔萨斯人口模型预测误差最大, logistic 模型比线性回归模型及马尔萨斯模型有很大的改善。**结论** logistic 模型更适合长期的人口预测, 误差更小, 模拟精度更好, 所预测的结果可信度更高。

关键词: 人口增长率; 马尔萨斯模型; logistic 模型; 线性回归模型; 人口预测

中图分类号: C 924.2

文献标志码: A

文章编号: 2096-3610 (2023) 06-0623-05

Population forecast nationwide based on linear regression model, Malthus Population Growth Model and logistic model

ZHAO Tian-wei, CHEN Hui-da* (School of Public Health, Guangdong Medical University, Dongguan 523808, China)

Abstract: **Objective** To have a better understanding of the change laws of Chinese population and predict the future population of China. **Methods** The population growth rate was calculated and the law of population growth was analyzed according to the population data from 1949 to 2019 published on the official website of the National Bureau of Statistics. The linear regression model, Malthusian population model and logistic model were established by using the population data from 1949 to 2014, which were evaluated with relative errors to predict the population in 2020–2040, and analyze the change trend of future population development. **Results** The population growth of China showed a downward trend, with negative population growth in 1960 and 1961, and positive growth in the rest years. It was found through model fitting and the comparison that the linear regression model had large errors, the Malthus population model had the largest prediction error, and the logistic model was much better than the linear regression model and Malthus model. **Conclusion** Logistic model is more suitable for long-term population prediction, with smaller error, better simulation accuracy and more reliable prediction results.

Key words: population growth; Malthus model; logistic model; linear regression model; population prediction

对于人口增长的预测, 已有许多学者进行了不同层面多维度的研究^[1-5]。为了更好地认识我国人口的数量变化规律, 并对我国未来人口进行预测, 本文利用新中国成立至 2019 年的数据, 建立不同的数学模型预测未来人口走向, 分别建立了线性回归模型、马尔萨斯人口增长模型和 logistic 模型, 确定我国的人口增长率, 并对 3 个模型进行数据检验和优劣对比。利用 1949–2014

年我国的人口数据建立线性回归模型、马尔萨斯人口增长模型和 logistic 模型, 然后利用 2015–2019 年我国实际人口数据与 3 个模型的拟合数据进行对比。最后利用 3 个模型预测我国 2020–2040 年的总人口数。

1 我国人口数据的描述性分析

根据我国 1949–2019 年的人口数据绘制线图可

收稿日期: 2022-11-26

基金项目: 湛江市非资助科技攻关计划项目 (2020B01302), 广东医科大学科研基金 (GDMUM2020038), 广东医科大学本科教学质量与教学改革工程项目 (1JG22158), 广东医科大学高等教育教学研究课题 (2JY22045), 广东医科大学“百项青年研究项目” (GDMUD2022024)

作者简介: 赵天伟 (1998–), 男, 本科, E-mail: 1789675620@qq.com

通信作者: 陈惠达 (1993–), 女, 硕士研究生, 讲师, E-mail: chenhdada@163.com

知(如图1),我国总人口数大体成正增长趋势,1960、1961年出现人口负增长(如图2),其余年份均为正增长。若从1949年开始(除去2019年),以10a为1个周期,我国的每10年人口增长率分别是21.83%、16.85%、19.32%、13.82%、10.70%、5.58%、4.56%,我国人口增长率呈下降趋势。

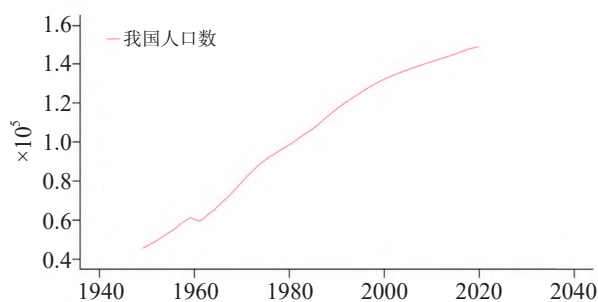


图1 1949-2019年我国人口数趋势图

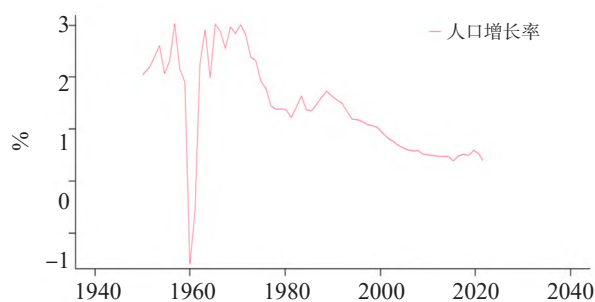


图2 1949-2019年我国人口增长率变动趋势图

2 模型的建立

2.1 线性回归模型的建立

作出如下假设:

假设1: 已得的人口数据真实、准确。

确定回归系数 a 。

利用matlab可以得到线性回归模型的拟合数据与实际数据的对比结果(如图3所示),其中蓝色实线表示线性回归模型拟合的我国人口数据,红色实线表示我国实际的人口数据。

该线性回归模型为:

$$y = 1387.21x + 2649764.19.$$

2.2 马尔萨斯模型的建立

排除干扰因素的影响,建立马尔萨斯模型;

作出如下假设:

假设1: 外界环境对我国人口数量的变化无干扰。

假设2: 人口增长率为 r ,时刻 t 的人口为 $x(t)$,初始时刻 $t=0$ 时,人口数为 x_0 。

假设3: 模型中,单位时间人口增长率为常数 r , t

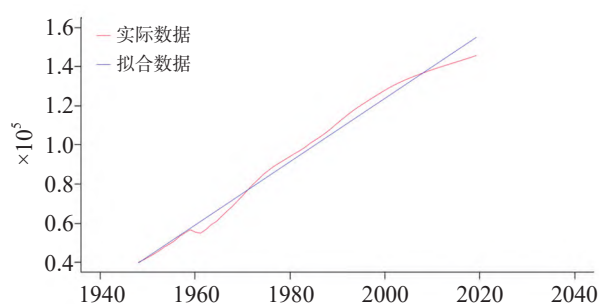


图3 线性回归模型拟合数据与实际数据的对比结果

时刻人口数量为连续、可微函数 $x(t)$ 。

假设4: 已得的人口数据真实、准确。

单位时间内 $x(t)$ 的增量为 $rx(t)$, $dx/dt=rx$, $x(0)=x_0$, 即 $x(t)=x_0e^{rt}$, 根据实际数据, 利用线性最小二乘法直接进行参数估计比较麻烦, 因此两边取对数, $\log x = \log x_0 + rt$, 原方程化为 $y=rt+a$ 。利用1949年($t=0$)至2014年我国人口数据进行拟合。

利用matlab可以得到拟合数据与实际数据的对比结果(如图4所示), 其中蓝色实线表示马尔萨斯人口模型拟合的我国人口数据, 红色实线表示我国实际的人口数据。

该马尔萨斯模型为: $y = 1.49 \times 10^{-2} \log(x) - 18.05$ 。

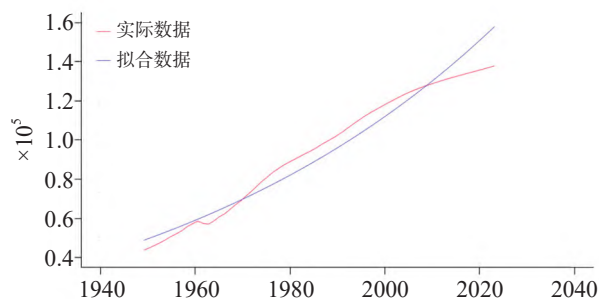


图4 马尔萨斯人口模型拟合数据与实际数据的对比结果

2.3 logistic模型的建立

作出如下假设:

假设1: 人口增长率为 r , 时刻 t 的人口为 $x(t)$, 初始时刻 $t=0$ 时, 人口数为 x_0 。

假设2: 函数 $x(t)$ 连续且可微。

假设3: 已得的人口数据真实、准确。

假设4: $r(x)$ 是 x 的线性函数。

单位时间内人口增量 $x(t+\Delta t) - x(t) = rx(t)\Delta t$, 令 $\Delta t \rightarrow 0$, 则可得 $x(t)$ 满足如下的微分方程 $dx/dt=rx$, $x(0)=x_0$ 。因为 r 随着人口数量 x 的增加而下降, 易得 $r(x)$ 为减函数, 则有 $dx/dt=r(x)x$, $x(0)=x_0$ 。 $r(x)$ 是 x 的线性函数, 即 $r(x) = r - sx$ ($r > 0, s > 0$)。此时引入最大人口环境容纳量 x_m , 当 $x=x_m$ 时, 人口数量

达到最大,此时人口增长率为 0,即 $r(xm) = r - sxm = 0$, 所以得到 $s = r/xm$, 于是由 $r(x) = r - sx$ ($r > 0, s > 0$) 得到如下的 logistic 模型 $r(x) = r(1 - x/xm)$, $x(0) = x_0$ 。

利用 matlab 可以得到拟合数据与实际数据的对比结果(如图 5 所示),其中蓝色实线表示 logistic 模型拟合的我国人口数据,红色实线表示我国实际的人口数据。

该 logistic 模型为:

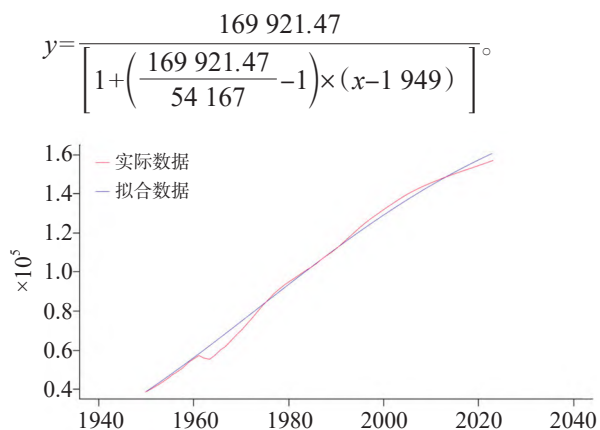


图 5 logistic 模型拟合数据与实际数据的对比结果

3 模型的评价

3.1 线性回归模型的评价

线性回归模型对 2015–2019 年拟合人口数与实际人口数对比见表 1。线性回归模型拟合 2015–2019 年的人口相对误差分别为 5.85%、6.24%、6.61%、7.21%、7.85%,平均相对误差为 6.75%。通过这 5 个已知的实际人口数据可以看出,利用线性回归模型拟合的人口数与实际人口数的相对误差均在 5% 以外,该拟合结果与实际人口数据存在出入,故改用马尔萨斯人口模型进行拟合预测。

表 1 2015–2019 年线性回归模型拟合人口总数与实际人口数对比 (万人)

年份	拟合人口数	实际人口数
2015	145 500	137 462
2016	146 900	138 271
2017	148 200	139 008
2018	149 600	139 538
2019	151 000	140 005

3.2 马尔萨斯回归模型的评价

马尔萨斯人口模型 2015–2019 年拟合人口数与实际人口数对比见表 2。马尔萨斯人口模型拟合 2015–2019 年人口的相对误差分别为 14.00%、15.06%、16.11%、17.46%、18.78%, 5 a 平均相对误差

为 16.28%。通过这 5 个已知的实际人口数据可以看出,利用马尔萨斯人口增长模型拟合的人口数与实际人口数相对误差均在 5% 以外,存在较大误差,因此我们改用 logistic 模型再次拟合预测我国人口数。

表 2 2015–2019 年马尔萨斯人口模型拟合人口数与实际人口数的对比 (万人)

年份	拟合人口数	实际人口数
2015	156 700	137 462
2016	159 100	138 271
2017	161 400	139 008
2018	163 900	139 538
2019	166 300	140 005

3.3 logistic 模型的评价

logistic 模型 2015–2019 年拟合人口数与我国实际人口数对比见表 3。logistic 模型拟合 2015–2019 年的人口的相对误差分别为 2.06%、2.05%、2.15%、2.34%、2.57%,平均相对误差为 2.23%,利用 logistic 模型拟合的人口数与实际人口数的相对误差均在 5% 以内,因此 logistic 模型的拟合结果可以接受。

表 3 2015–2019 年 logistic 模型拟合人口总数与实际人口数的对比 (万人)

年份	拟合人口数	实际人口数
2015	140 300	137 462
2016	141 100	138 271
2017	142 000	139 008
2018	142 800	139 538
2019	143 600	140 005

3.4 3 种模型拟合效果对比

如图 6 所示,红色实线为我国实际人口数据,黄色实线表示 logistic 模型拟合人口数据,蓝色实线表示马尔萨斯人口模型拟合的我国人口数据,绿色实线表示线性回归模型拟合的人口数据。

通过与这 5 个已知的实际人口数据比较,由图 7 可见,利用 logistic 模型拟合的人口数与实际人口数的相对误差均在 5% 以内,因此 logistic 模型的拟合结果可以接受。

4 2020–2040 年人口预测

通过建立线性回归模型、马尔萨斯人口模型及 logistic 模型预测 2020–2040 年的人口总数见表 4。用此 3 种模型拟合的人口数分别与 2015–2019 年实际人口数进行对比,以及对我国 2040 年人口数进行预测,可发现以下 3 点:(1)线性回归模型和马尔萨斯人口模型并不能满足我们对我国未来人口预测的精度要

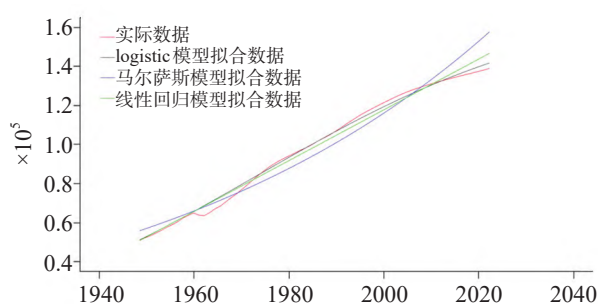


图6 3种模型的拟合效果对比图

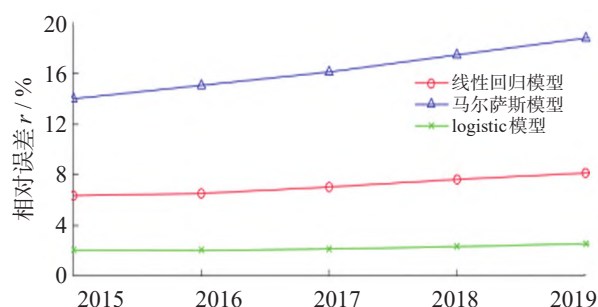


图7 2015–2019年3种模型的相对误差对比

求,存在误差比较大;而logistic模型能很好地满足对我国未来人口预测的精度要求,存在的误差较小,有较高的可信度。(2)3个人口预测模型相比(见图8),马尔萨斯人口模型相对误差最大,其次是线性回归模型,logistic模型的相对误差最小,原因在于本文预测的时间较长,研究的年份较多,所以相比之下logistic模型可以获得良好的预测效果。(3)到2040年我国总人口规模,线性回归模型预测达到180 100万人,马尔萨斯人口模型预测达到227 400万人;logistic模型预测达到156 200万人。这是3种模型仅从历史数据的变化趋势所做出的3个预测(如图8),未考虑国家放开三孩等政策因素,这些因素会对我国未来实际的人口有指导作用,同时对我国实际人口规模产生巨大影响。

综合以上3点,这能较好地说明logistic模型更适合长期的人口预测。而本文目的是预测我国2040年人口数量规模,选用logistic模型误差更小,模拟精度更好,所预测的结果可信度更高。

5 结语

自1949年新中国成立以来,我国人口变化的历史帮助我们加深了对人口发展规律的理解,同时也为我国的发展提供了一些基本经验,其中包括在发展进程中取得的成就和从中汲取的教训。这些成就和教训对促进我国未来的发展极为宝贵,有着不同的积极

表4 3种模型预测2020–2040年的人口总数 (万人)

年份	线性回归模型预测人口数	马尔萨斯模型预测人口数	logistic模型预测人口数
2020	152 400	168 800	144 300
2021	153 800	171 400	145 100
2022	155 200	173 900	145 800
2023	156 600	176 500	146 500
2024	158 000	179 200	147 200
2025	159 300	181 900	147 900
2026	160 700	184 600	148 600
2027	162 100	187 400	149 200
2028	163 500	190 200	149 800
2029	164 900	193 000	150 500
2030	166 300	195 900	151 100
2031	167 700	198 900	151 600
2032	169 000	201 900	152 200
2033	170 400	204 900	152 800
2034	171 800	208 000	153 300
2035	173 200	211 100	153 800
2036	174 600	214 300	154 300
2037	176 000	217 500	154 800
2038	177 400	220 700	155 300
2039	178 800	224 100	155 700
2040	180 100	227 400	156 200

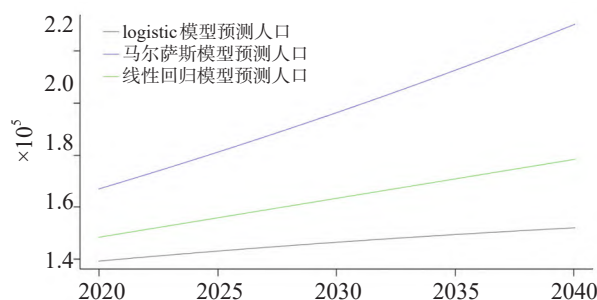


图8 3种模型的预测结果对比图

和消极的启迪。在中国新的社会主义建设的最初,我们认为人口越多,生产力越多就越有利于发展。在经历了20世纪60年代和70年代的曲折发展,我们还认为,人口数少有利于经济现代化。在进入21世纪之后,我们日益关切的是,人口减少和结构变化将对未来的经济发展产生消极影响。2010年,杨颖^[1]论述了数学模型在认识人口规律和预测未来人口数量中的应用。多位学者用不同模型^[2-3]预测了不同地区^[4-5]不同时间段的人口数量。我们认为,理论上线性回归模型是不理想的,因为人口增长的内在联系错综复杂,无法用1个变量很好的描述;对于马尔萨斯人口模型,在短期内对人口数进行预测是比较准确的。但在马尔萨斯模型中,人口数将随着时间的增长而无限增长,这

显然是不可能的,在考虑资源与环境等因素的干扰下, logistic 模型可能能够更好地预测未来我国人口数。

参考文献:

- [1] 杨颖. 数学建模在实践中的应用[J]. 长春师范学院学报(自然科学版), 2010, 29(6): 106-108.
[2] 刘令, 杨力, 林源. 基于指数增长模型的全国人口预测[J]. 山

东工业技术, 2014(16): 164.

- [3] 黄晨. 基于 logistic 模型对中国未来人口的预测分析[J]. 中国新通信, 2019, 21(23): 160.
[4] 刘亚平, 刘云忠. 美国人口的预测[J]. 经济研究导刊, 2020(3): 175-177.
[5] 高梅, 康宝生, 曹黎侠. 数学模型在西安市人口预测中的应用[J]. 西安工业大学学报, 2019, 39(4): 373-377.

~~~~~  
(上接第 622 页)

自研的核酸检测成本(16.0 元/人份)远低于凯杰(32.7 元/人份)和诺唯赞(20.0 元/人份),说明南芯自研核酸试剂性价比极高;而美基成本(15.0 元/人份)虽然最低,但其为 96 人份预分装板,样品起始量不足时易造成试剂浪费。

综上,南芯自研核酸提取试剂,其提取流程简单、操作方便,所提取的核酸纯度与市面凯杰、诺唯赞等 3 家品牌几乎一致,当菌液浓度 $\geq 10^5$  CFU/mL 时核酸提取浓度仅低于凯杰品牌,检测成本则低于凯杰和诺唯赞品牌,提取效能亦远超 3 家核酸提取试剂品牌,综合性价比最高。南芯自研核酸提取试剂不仅可满足临床检测需要,也适用于常规基础实验,值得推广。

#### 参考文献:

- [1] RANA J S, KHAN S S, LLOYD-JONES D M, et al. Changes in mortality in top 10 causes of death from 2011 to 2018[J]. J Gen Intern Med, 2021, 36(8): 2517-2518.  
[2] ZHU Y G, TANG X D, LU Y T, et al. Contemporary situation of community-acquired pneumonia in China: A systematic review[J]. J Transl Int Med, 2018, 6(1): 26-31.  
[3] SCHLABERG R, CHIU C Y, MILLER S, et al. Validation of metagenomic next-generation sequencing tests for universal pathogen detection[J]. Arch Pathol Lab Med, 2017, 141(6): 776-786.  
[4] 张淋, 洪城, 新科, 等. 宏基因组学第二代测序技术对比传统

实验室微生物培养在脓毒症病原学诊断中的优势[J]. 中国急救医学, 2022, 42(2): 114-120.

- [5] 江婷婷, 林彦锋. 人体内微生物游离 DNA 在病原微生物检测中的应用[J]. 安徽医科大学学报, 2023, 58(2): 327-331.  
[6] 李俊红. 病原微生物对人体健康的危害及检测[J]. 农业开发与装备, 2016(4): 43.  
[7] 刘瑾, 龙运峰, 罗勇, 等. 胃液病原微生物核酸检测在早产儿早期感染诊断中的价值[J]. 邵阳学院学报(自然科学版), 2022, 19(5): 13-18.  
[8] 甘明宇, 李刚, 俞惠, 等. 宏基因组 Nanopore 测序在儿童呼吸道病原微生物快速检测中的应用价值[J]. 中国循证儿科杂志, 2020, 15(5): 4.  
[9] 卢运战, 程帮照, 杨金兵. 新城疫病毒核酸提取方法的优化与比较[J]. 畜牧与饲料科学, 2017, 38(4): 11-13.  
[10] 马晓华, 柴萨萨, 利利, 等. 粪便和组织样本高通量测序前处理方法的评估[J]. 中华实验和临床病毒学杂志, 2022, 36(1): 103-110.  
[11] 杨焕蝶, 杨金玉, 陈相艳, 等. 肠道集聚性大肠杆菌基因组脱氧核糖核酸提取方法比较与改进[J]. 济南大学学报(自然科学版), 2021, 35(2): 144-149.  
[12] 刘燕萍, 刘业英, 文岚. 临床检验标本中病原微生物的 DNA 快速提取方法分析[J]. 医学信息, 2013(24): 85.  
[13] 王浩, 周俊宇, 苏锦华. 三种方法提取猪丹毒杆菌核酸的比较[J]. 今日畜牧兽医, 2019, 35(3): 3.  
[14] 刘晨曦, 姚延禄, 周新丽. 用于磁珠法核酸提取的微流控芯片及自动化平台[J/OL]. 分析试验室, 2023, DOI: 10.13595/j.cnki.issn1000-0720.2022.102402.