

基于灰色预测模型的陕西省常住人口预测研究

李劲彬^{1,2,3,4,5}, 花东文^{1,2,3,4,5}

¹陕西地建土地工程技术研究院有限责任公司, 陕西 西安

²陕西省土地工程建设集团有限责任公司, 陕西 西安

³自然资源部退化及未利用土地整治工程重点实验室, 陕西 西安

⁴陕西省土地整治工程技术研究中心, 陕西 西安

⁵自然资源部土地工程技术创新中心, 陕西 西安

收稿日期: 2022年4月15日; 录用日期: 2022年5月17日; 发布日期: 2022年5月25日

摘要

为了更好地预测分析陕西省常住人口, 本文利用国家统计局官网最新统计资料, 以2010~2019年陕西省常住人口数据为基础, 建立GM(1,1)灰色预测模型, 并用2020年常住人口数值对模型预测精度进行误差检验, 结果显示模型预测精度高达99.54%。因此, 使用该模型对2021~2025年陕西省常住人口数据进行预测研究, 结果表明: 2022年陕西省常住人口将超过4000万, 到2025年将达到4092.8万。整体来看, 未来5年陕西省常住人口数值以每年23.7万人的增速持续增长。

关键词

GM(1,1), 陕西省, 人口预测, 灰色预测, 预测精度

Prediction and Analysis of Resident Population in Shaanxi Province Based on GM(1,1) Grey Prediction Model

Jinbin Li^{1,2,3,4,5}, Dongwen Hua^{1,2,3,4,5}

¹Institute of Land Engineering and Technology, Shaanxi Provincial Land Engineering Construction Group Co., Ltd., Xi'an Shaanxi

²Shaanxi Provincial Land Engineering Construction Group Co., Ltd., Xi'an Shaanxi

³Key Laboratory of Degraded and Unused Land Consolidation Engineering, Ministry of Natural Resources, Xi'an Shaanxi

⁴Shaanxi Provincial Land Consolidation Engineering Technology Research Center, Xi'an Shaanxi

⁵Land Engineering Technology Innovation Center, Ministry of Natural Resources, Xi'an Shaanxi

Received: Apr. 15th, 2022; accepted: May 17th, 2022; published: May 25th, 2022

Abstract

In order to better predict and analyze the resident population of Shaanxi Province, this paper uses the latest statistical data on the official website of the National Bureau of statistics, based on the resident population data of Shaanxi Province from 2010 to 2019, establishes the GM (1,1) grey prediction model, and tests the prediction accuracy of the model with the resident population value in 2020. The results show that the prediction accuracy of the model is as high as 99.54%. Therefore, the model is used to predict the resident population data of Shaanxi Province from 2021 to 2025. The results show that the resident population of Shaanxi Province will exceed 40 million in 2022 and 40.928 million in 2025. Overall, the resident population of Shaanxi Province will continue to grow at an annual growth rate of 237000 in the next five years.

Keywords

GM(1,1), Shaanxi Province, Population Prediction, Grey Prediction, Prediction Accuracy

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

陕西省统计局最新发布的《陕西省第七次全国人口普查主要数据公报》显示,截止2020年11月1日,全省常住人口总数为3953万,其中居住在城镇的人口数量约为2477万,占常住人口的62.66%,与2010年陕西省第六次全国人口普查相比,全省常住人口总数增加218万,城镇人口增加771万人,人口比重提升16.96%。陕西省地处西北内陆腹地,是连接中国西北、西南、东部以及中部的重要枢纽。常住人口可以客观、真实地反映出陕西省的经济和社会发展水平,有助于区域发展规划以及国民经济计划的制定[1]。通过对常住人口的统计,可以计算出陕西省的人均生产总值,即人均GDP,其是衡量省内人民生活水平的重要参考依据,对于把握陕西省经济的发展运行状况意义重大。

灰色系统理论是由我国著名控制论学者邓聚龙教授在二十世纪八十年代首次提出的一种不确定性系统预测理论,通过对部分已知信息进行序列生成、建模来寻找有价值的信息,进而实现对系统运行以及演化规律的有效掌控[2][3]。GM(1,1)预测模型是灰色理论中的核心模型,适合对数据量少、信息贫乏的研究对象进行中短期预测,因此被广泛应用于经济、农业、工业、水文等领域。杨伟传等通过应用灰色系统建模方法预测了广东省江门市GDP,并对预测数据进行了比较,结果显示模型预测精度很高[4]。吴鹏等通过灰色系统关联度分析理论,应用优化后的多维灰色GM(1,N)模型,模拟预测了中国GDP的发展,结果发现优化后的模型相比传统模型拟合精度更高[5]。吴华安等应用多维灰色系统预测模型OGM(1,N)对重庆市人口密度进行了模拟和预测,并取得了良好的效果,对城市人口相关政策的制定具有积极的意义[6]。

2. 分析方法简介

2.1. GM(1,1)预测模型的构建

设序列 $X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$ 为某一预测对象的原始数列, 并且 $x^{(0)}(k) \geq 0, k = 1, 2, \dots, n$; 为弱化序列波动性及随机性, 对 $X^{(0)}$ 进行一次累加, 生成序列 $X^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n))$, 其中 $x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i), k = 1, 2, \dots, n$ 。

设 $Z^{(1)} = (z^{(1)}(2), z^{(1)}(3), \dots, z^{(1)}(n))$ 为 $X^{(0)}$ 的紧邻均值生成序列, 其中 $z^{(1)}(k) = \frac{1}{2}(x^{(1)}(k) + x^{(1)}(k-1)), k = 2, 3, \dots, n$, 则 $x^{(0)}(k) + az^{(1)}(k) = b$ 为 GM(1,1)的灰微分方程模型, 其对应的微分方程为 $\frac{dX^{(1)}}{dt} + aX^{(1)} = b$, 也称 GM(1,1)的白化型, 其中 a, b 均为特定参数, a 为系统发展系数, b 为灰作用量, 求解该微分方程得时间响应函数: $\hat{x}^{(1)}(k+1) = \left(x^{(0)}(1) - \frac{b}{a}\right)e^{-ak} + \frac{b}{a}, k = 1, 2, \dots, n$ 。 $X^{(0)}$ 的预测值可通过公式 $\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) = (1-e^a)\left(x^{(0)}(1) - \frac{b}{a}\right)e^{-ak}$ 计算得到。

$$\text{令 } Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}, \quad B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix}, \quad u = \begin{bmatrix} a \\ b \end{bmatrix}, \text{ 则有方程 } Y = Bu, \text{ 应用最小二乘法求解方程, 得特}$$

定参数: $u = (a, b)^T = (B^T B)^{-1} B^T Y$ 。

2.2. 预测模型精度检验

为评价预测模型 GM(1,1)的精度和效果, 需对所建模型的精度进行检验, 只有模型检验合格后才能用于相关预测, 预测结果才有意义。精度检验通常采用残差检验法、后验差检验法以及关联度检验法三种方法。本文将采用后验差检验法对模型精度进行检验, 具体方法如下:

分别计算原始序列的均值 $\bar{x}^{(0)} = \frac{1}{n} \sum_{k=1}^n x^{(0)}(k)$ 、标准差 $S_1 = \sqrt{\frac{\sum [x^{(0)}(k) - \bar{x}^{(0)}]^2}{n-1}}$ 以及残差序列的均值 $\Delta^{(0)} = \frac{1}{n} \sum_{k=1}^n [x^{(0)}(k) - \hat{x}^{(0)}(k)]$ 和绝对误差序列的标准差 $S_2 = \sqrt{\frac{\sum [\Delta^{(0)}(k) - \Delta^{(0)}]^2}{n-1}}$; 进而计算出均方差比 $C = \frac{S_2}{S_1}$, 小误差概率 $P = p\{|\Delta^{(0)}(k) - \Delta^{(0)}| < 0.6745S_1\}$, 然后参照表 1 标准进行预测模型的精度等级划分, 如果某预测模型 GM(1,1)的 C, P 值不在同一列, 则精度等级就低不就高, 如按照 C 其精度为 1 级, 而按 P 其精度为 2 级, 则该模型的预测精度为 2 级。

Table 1. Model prediction accuracy rating table

表 1. 模型预测精确度等级评价表

精度等级	1 级(好)	2 级(合格)	3 级(勉强合格)	4 级(不合格)
均方差比 C	$C \leq 0.35$	$0.35 < C \leq 0.5$	$0.50 < C \leq 0.65$	$C > 0.65$
小误差概率 P	$P \geq 0.95$	$0.80 \leq P < 0.95$	$0.70 \leq P < 0.80$	$P < 0.70$

3. 陕西省常住人口 GM(1,1)预测模型的建立

本文根据陕西省 2010~2019 年常住人口数据建立 GM(1,1)预测模型, 并预留 2020 年常住人口数据对所建预测模型的精度进行误差检验。所用数据来自国家统计局官网

(<https://data.stats.gov.cn/easyquery.htm?cn=E0103>), 具体常住人口数据如表 2 所示。

Table 2. 2010~2020 resident population data in Shaanxi Province (unit: ten thousand people)

表 2. 陕西省 2010~2020 年常住人口数据(单位: 万人)

年份	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
常住人口	3735	3765	3787	3804	3827	3846	3874	3904	3931	3944	3953

数据来源: 国家统计局官网(2010~2020 年)。

通过对原数据序列 $X^{(0)}$ 进行准光滑性检验、准指数规律检验以及级比检验发现, 当 $k \geq 4$ 时, 这些数据的光滑比 $\rho(k) = \frac{x^{(0)}(k)}{x^{(0)}(k-1)}$ 是 k 的递减函数, 且 $\rho(k) < 0.5$; 当 $k \geq 4$ 时,

$\sigma^{(1)}(k) = \frac{x^{(1)}(k)}{x^{(1)}(k-1)} \in [1, 1 + \delta]$ ($\delta = 0.5$), 准指数规律满足; 级比 $\sigma^{(0)}(k) = \frac{x^{(0)}(k-1)}{x^{(0)}(k)}$ 均属于

$\left(e^{-\frac{2}{n+1}}, e^{\frac{2}{n+1}} \right) = (0.8338, 1.1994)$, 因此数据通过了准光滑性检验、准指数规律检验和级比检验, 可以对原始数据直接进行 GM(1,1)预测分析。

根据 GM(1,1)模型原理及其构建方法, 经计算得到陕西省常住人口预测模型为:

$$\hat{x}^{(1)}(k+1) = 621200e^{0.006k} - 617470, k = 1, 2, \dots, n$$

经后验差检验得: $C = 0.004$, $P = 1$, 因此对比表 1 可得, 模型预测精度等级为 1 级, 预测模型合格。残差具体计算过程见表 3。

Table 3. Residual test table

表 3. 残差检验表

k	模型生成值 $\hat{x}^{(1)}(k)$	实际值 $x^{(1)}(k)$	还原数据 $\hat{x}^{(0)}(k)$	实际数据 $x^{(0)}(k)$	残差 $\Delta^{(0)}(k)$
1	3735	3735	3735	3735	0
2	7496.15	7500	3761.15	3765	3.85
3	11280.08	11287	3783.93	3787	3.07
4	15086.92	15091	3806.84	3804	-2.84
5	18916.81	18918	3829.89	3827	-2.89
6	22769.88	22764	3853.08	3846	-7.08
7	26646.29	26638	3876.40	3874	-2.4
8	30546.16	30542	3899.87	3904	4.13
9	34469.65	34473	3923.49	3931	7.51
10	38416.89	38417	3947.24	3944	-3.24

4. 基于 GM(1,1)预测模型的未来 5a 人口预测

Table 4. Predicted value of permanent resident population in Shaanxi Province from 2021 to 2025

表 4. 陕西省 2021~2025 年常住人口预测值

年份	预测值(万人)	人口年净增长值(万人)	增长率(%)
2020	3971.14	-	-
2021	3995.18	24.04	6.05
2022	4019.37	24.19	6.05
2023	4043.71	24.34	6.06
2024	4068.19	24.48	6.05
2025	4092.82	24.63	6.05

由表 4 可知, 2020 年陕西省常住人口预测值为 3971.1 万人, 与实际人口 3953 万人相比, 相对误差仅为 0.46%, 远远小于 5%, 预测精度高达 99.54%。未来 5a (2021~2025 年)陕西省常住人口数量在没有自然灾害以及政策因素影响下, 会连续稳步增长, 增长幅度与前几年基本相当, 年平均增长率基本维持在 6.05%。

5. 结论

灰色预测主要包含时间序列、灾变、拓扑以及系统等四方面预测, 其主要特点是少数据、有因果和具有可检验性。由于灰色系统分析并非对所有系统的相关数据都可以进行预测分析, 因此可检验性就十分必要。因此本文在分析预测前对常住人口数据进行了准光滑性检验、准指数规律检验以及级比检验, 以确保数据符合 GM(1,1)预测模型建模要求, 再通过对所建模型精度的检验, 最大程度提高了分析结果的可靠性。从预测数据来看, 陕西省常住人口未来 5 年会持续增长, 且增速约为每年 24.3 万人。

参考文献

- [1] 彭洁. 浅析常住人口统计工作的意义和方法[J]. 中国科技投资, 2013(34): 443.
- [2] 孔雪, 王丽, 冯益华. 灰色预测 GM(1,1)模型应用现状与展望[J]. 齐鲁工业大学学报, 2018, 32(6): 49-53.
- [3] 徐宁, 丁松, 公彦德. 灰色 GM(1,1)预测模型及拓展研究进展[J]. 数学的实践与认识, 2021, 51(13): 52-59.
- [4] 杨伟传, 谭宇柱. GM(1,1)模型在经济分析中的应用[J]. 五邑大学学报(自然科学版), 2008, 22(3): 66-69.
- [5] 吴鹏, 邱赛兵. 基于优化多维灰色模型的宏观经济发展预测[J]. 统计与决策, 2020, 36(3): 42-45.
- [6] 吴华安, 曾波, 彭友, 周猛. 基于多维灰色系统模型的城市人口密度预测[J]. 统计与信息论坛, 2018, 33(8): 60-67.