

基于回归分析的西部生育率影响因素分析

张 林

广西师范大学数学与统计学院, 广西 桂林

收稿日期: 2023年3月24日; 录用日期: 2023年4月14日; 发布日期: 2023年4月27日

摘 要

本文将通过多元线性回归模型对相关数据进行分析, 利用R软件, 将收集到的15个指标与生育率建立模型, 通过相关模型诊断和AIC模型优化找出对西部生育率影响最大的几个因素。由分析结果可知, 影响西部生育率的因素包括: 结婚人数、中专及大学、农村人均可支配收入、城镇人均可支配收入、男性适婚人数、女性适婚人数、受教育程度, 并提出了相对应的建议。

关键词

多元线性回归模型, 生育率, AIC模型优化

Based on Analysis of Regression Analysis in the West of Fertility Factors

Lin Zhang

School of Mathematics and Statistics, Guangxi Normal University, Guilin Guangxi

Received: Mar. 24th, 2023; accepted: Apr. 14th, 2023; published: Apr. 27th, 2023

Abstract

This paper will analyze the relevant data by multiple linear regression model, and use R software to model the 15 indicators collected with the fertility rate, and find out the factors that have the greatest influence on the fertility rate in the west by correlation model diagnosis and AIC model optimization. From the analysis results, it can be seen that the factors affecting the fertility rate in the west include: the number of marriages, secondary school and university, rural per capita disposable income, urban per capita disposable income, the number of marriageable men, the number of marriageable women, and the education level, and the corresponding suggestions are made.

Keywords

Multiple Linear Regression Model, Fertility, AIC Model Optimization

Copyright © 2023 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

我国西部地区的人口密度一直偏低，一部分原因来自它独特的地形地貌，还有一部分原因来自不好的气候条件。长此以往，西部地区的人口增长就成为一项难题。随着改革开放的到来，我国开始逐渐开放二胎政策[1]。伴随着中国经济腾飞的十年，也出现了许多新的问题需要解决，比如性别比例失调、人口老龄化、环境污染、劳动力规模持续萎缩等问题[2]。随着人们受教育的程度越高，对实现自身价值的渴望就越强，越来越多的青年人把时间和精力投身于事业之中，忽略了自身的情感，这也是我国生育率持续下降的重要原因。为了针对性的解决人口的问题，我国在2016年开放了二胎政策，政策实施初期取得一些成效。2016年的新生人口数比前一年增加了131万，但随即生育率却快速下降，新生儿出生率下降至自建国以来的最低值，仅为10.94% [3]。根据《中国统计年鉴2021》显示2020年全国人口出生率仅为0.85%，首次跌破了1%，创造了43年来的新低，若持续发展下去，未来我国极大可能出现总人口负增长的情况。2021年5月31日，中央政治局会议审议通过《关于优化生育政策促进人口长期均衡发展的决定》，提出“进一步优化生育政策，实施一对夫妻可以生育三个子女政策及配套支持措施” [4]。

2. 文献综述

许多学者对生育率进行了不同的研究，有些学者将生育率与育龄人群的生育意愿结合历史因素进行了论述，有些学者研究了人口流动对生育水平的影响[5]，张孝栋和张雅璐[6]指出无论是宏观还是微观的个体经济发展水平亦或是经济负担的加重都会对使生育率下降；穆光宗[7]建议，为提升生育率最好在婚姻中的责任理论能够回归传统；曹园[8]则指出养老保险在一定程度上可以促进生育率的增加；李飏[9]通过相关数据分析得出了互联网对女性的生育率存在一定的影响，伴随着互联网的出现，增加了女性的就业，例如各大平台的直播等，更加便于女性的就业与选择，增加了女性的收入和更好的提升自己，但生育率有所下降；倪云松[10]的研究发现，房价或持有的房产无论是从哪一方面来看都会对女性的生育带来一定的影响，而且常常带来的影响都是负面的；何林浩与陈梦[11]利用排序logit模型来验证了在已结婚的夫妻中，当女性的受教育程度相对男性较高时，则生育率越低；张永英[12]研究表明，可以通过保障就业的合法权益来提升妇女生育意愿；杨成钢和孙晓海[13]表明，对于生育率降低的风险，需要尽量做到风险可以减缓以及能够适应风险等；毕灵斐[14]指出，通过结合各类妇女保障政策能在一定程度上对生育率产生积极的影响；张豫南[15]表示，在社会的相关保障等各方面的考虑下，女性则更加可能推迟结婚，教育水平越高、独立性越强，越自主的女性，进入婚姻的时间会相对更晚一点。

3. 实证分析

3.1. 变量选择

生育率的相关概念：一般生育率指的是将育龄妇女简单划分为一个整体的类别，再考虑生育率。对

于分年龄生育率来说, 年龄段不同, 生育率也会相差。例如, 20~29 岁正处于青年时期, 正是处于最好的育龄阶段, 显然此时的生育率处于最高值, 但是另一年龄段的生育率很高低, 发生了很大转变, 年龄很小或者很大生育率相应就比较小。

综合来看, 由于总和生育率的数据并没有进行统计, 因此选则计算一般生育率来衡量生育率水平是一个不错的选择, 每 1000 名妇女的全年活产婴儿数。

3.2. 模型假设

x_1, x_2, \dots, x_p 多元线性回归模型中存在多个自变量。设 Y 为因变量, x_i 为自变量, ε 为随机误差。多元线性回归模型的一般形式如下:

$$y = f(x_1, x_2, \dots, x_p) + \varepsilon \quad (1)$$

解决实际问题时, 如获得 n 组数据 $(x_{i1}, x_{i2}, \dots, x_{ip}, y_i) (i = 1, 2, \dots, n)$, 可用以下方程对回归方程进行表示:

$$\begin{cases} y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \dots + \beta_p x_{1p} + \varepsilon_1 \\ y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \dots + \beta_p x_{2p} + \varepsilon_2 \\ \dots \\ y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \dots + \beta_p x_{np} + \varepsilon_n \end{cases} \quad (2)$$

也可用矩阵形式表示为:

$$y = X\beta + \varepsilon \quad (3)$$

式中, β_0 为回归常数, $\beta_1, \beta_2, \dots, \beta_p$ 为回归系数, 求回归参数 β_0 和 $\beta_1, \beta_2, \dots, \beta_p$ 的方法有最小二乘法、最大似然法等。

3.3. 实证分析过程

本文选取西部 12 个省市及自治区进行研究, 选取的年份区间为 2001~2020 年。本文所取数据, 如果没有特殊说明, 都是通过各大统计网站上将数据进行收集, 主要有中国知网、中国统计年鉴、中国劳动统计年鉴、历年的民政统计年鉴、各省对应的政府网站等。

3.3.1. 描述性统计

将得到的数据进行汇总。首先, 将当年某省出生婴儿总数 - 当年某省出生婴儿总数 \times 当年全国出生婴儿死亡率得到的结果为生育人数, 我们在这里将视为某省当年的生育率; 其次, 将使用上年年末总人口数和本年年末总人口数的平均数(抽样估计)作为今年人口总数的估计全国适婚男性 = 总人口 \times 适婚男性占比, 使用插值法填补缺失年份数据计算各省每年受调查男性在全国受调查男性中占的比例, 各省适婚男性 = 全国适婚男性 \times 各省男性占比, 各省适婚女性 = 全国适婚女性 \times 各省女性占比。第三, 将消费价格指数和商品零售指数都换算成相对上一年的比。第四, 运用均值插补法对缺失数据进行插补, 有效的处理了缺失值异常值等问题。

影响生育率的因素有很多, 生育率上升或是下降是一个复杂的过程, 是经济、政治、社会、文化以及心理等很多因素共同作用的结果。微观因素形如家庭收入情况、家庭经济为什么类型、家庭消费的高或低模式以及日常生活中的生活质量等, 也有经济发展、教育、卫生医疗等因素。在以往研究的论文基础上, 初步推测影响生育率的因素后, 确定要收集的数据指标, 分别表示结婚人数(万对)、离婚人数(万对)、未上过学、小学(%)、初中(%)、高中和中等职业教育(%)、中专及大学(%)、房价(元)、农村人均可支配收入(元)、城镇人均可支配收入(元)、男性适婚人数(万)、女性适婚人数(万)、出生率(%)、增长率(%)。

3.3.2. 多重相关系数和变异系数

变量之间存在多重共线性。比如：生育率和结婚人数呈显正相关，房价与农村人均可支配收入和城镇人均可支配收入呈正相关等。总的来说，自变量与生育率都有或强或弱的相关性，因此，考虑以生育率为因变量，其余变量为自变量进行多元线性回归分析。

通过分析知，自变量中只有结婚人数、中专及大学、农村人均可支配收入、城镇人均可支配收入、男性适婚人数、女性适婚人数、男女适婚比例的系数通过了 t 检验，而其他的自变量的均未通过；再看该模型的 F 检验，结果中的 P 值远远小于 0.05，可决系数和调节可决系数都是在 0.5 附近，综合说明该模型可以解释数据变异的 50%左右，采用多元线性回归模型是合适的。

3.3.3. 主成分回归

通过前面的分析已经发现，多个自变量之间存在较强的相关性，因此可以对这些自变量计算主成分。我们使用 `prcomp` 函数对训练集的自变量计算 PCA (主成分分析)，见图 1 所示。由分析可知，第一主成分解释了 84.8%，第二主成分解释了 14.03%，第三主成分解释了 0.8%，第四主成分解释了 0.3%，我们选择 2 个主成分来解释。

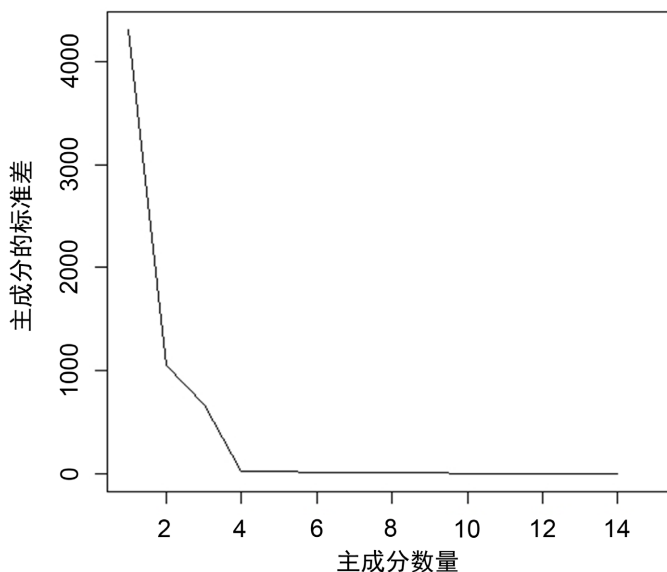


Figure 1. Plot of principal components

图 1. 主成分

3.3.4. AIC 模型优化

采用 AIC 准则即最小化信息量准则进行最优模型选择，加权拟合精度和参数个数，模型优化后得出以下七个变量对生育率影响较为明显，分别为结婚人数(X_1)、中专及大学(X_2)、农村人均可支配收入(X_3)、城镇人均可支配收入(X_4)、男性适婚人数(X_5)、女性适婚人数(X_6)、未上学(X_7)、出生率(X_8)、初中(X_9)。

$$Y = -148.1 + 0.7720X_1 - 1.677X_2 + 0.0029X_3 - 0.0009X_4 - 0.1557X_5 \\ + 0.1632X_6 - 0.3494X_7 + 1.355X_8 + 0.4279X_9$$

3.4.5. 统计性检验

见表 1 中显示了该模型的标准误差、t 值、p 值，回归系数检验的 t 统计量观测值和相应的概率 p 值。由表中的系数可知，利用 AIC 模型优化后的变量对生育率都有显著的影响。

Table 1. Sample correlation matrix

表 1. 样本相关矩阵

模型	可决系数: 0.6934	标准误差	t 值	Pr (> t)
	可决系数: 0.68			
常量		3.578e+01	-4.138	4.93e-05
结婚人数		7.720e-01	10.054	2e-16
中专及大学		-1.677	-3.179	0.001684
农村人均可支配收入		2.995e-03	2.998	0.003016
城镇人均可支配收入		-9.183e-04	3.588e-04	0.011135
男性适婚人数		-1.557e-01	-2.896	0.004143
女性适婚人数		1.632e-01	2.892	0.004196
未上学		-3.494e-01	-2.691	0.007661
出生率		1.355e+00	3.410	0.000769
初中		4.279e-01	2.785	0.005810

3.4.6. 未来十年预测

由图 2 知, 未来十年的生育率在 39%左右, 对比 2020 年的最低生育率为 42.85%下降了大约 3%, 这是一个非常严峻的趋势, 有相对性做出相应措施迫在眉睫。

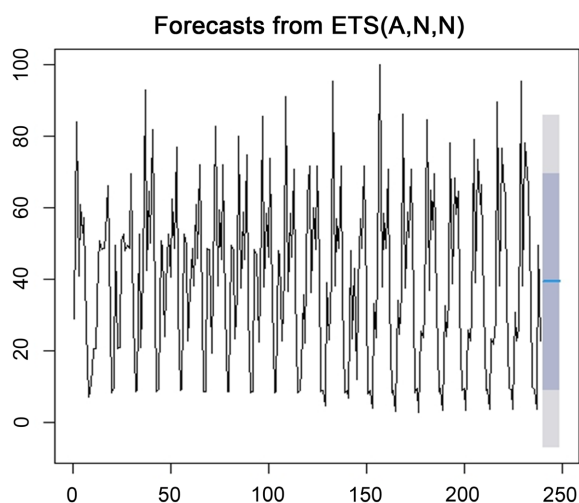


Figure 2. Forecast figure

图 2. 预测图

4. 结论及对策

4.1. 主要结论

通过上述分析过程, 我们通过查找文献和初步判断确定了 15 个变量, 对选择的 15 个对西部出生人口产生潜在影响的变量进行建模, 从分析结果来看, 发现 15 个变量中有 9 个变量对生育率产生了显著影响, 分别是: 结婚人数、出生率、男性适婚人数、女性适婚人数、中专及大学、农村人均可支配收入、城镇人均可支配收入、未上学、初中。无论是 2020 年一季度的婚姻登记对数大幅度的下降还是人口出生

数量的骤降,都表明提高结婚率是很有必要的,而结合实际知道结婚率的增加也会受很多因素影响,可能女性的独立意识的增加,或是婚姻成本的增加致使男性在面对婚姻问题时打退堂鼓,又或是其他的原因,我们有不一样的看法,或许在政策上国家可以将男性适婚年龄和女性适婚年龄降低一下,可能会在一定程度上提高生育率。但本文的研究仍然存在一定的局限性,如考虑的变量相对主观,没有考虑可能降低男性和女性的适婚年龄可能会给社会的发展带来什么样的影响,以及这个政策的实施是否可行等各方面的因素可能都会带来一定的影响。

4.2. 可行性建议

针对结论提出以下两点建议:

1) 把生育政策落到实处。

当今社会的生育成本太高,大部分年轻人的经济压力过大,工作节奏太快等原因,从而忽略了个人的感情问题,大部分年轻人奋斗事业而选择晚婚晚育,这也是我们国家经济发展与人口发展不平衡的原因所在,因此,政府要把鼓励生育的政策落到实处。比如,减少九九六的工作时间,增加国家法定节假日的休息时间,以及增加员工福利等宏观调控。

2) 树立正确的婚恋观。

针对西部地区的孩子,存在这两种极端的婚恋观。及早辍学的孩子会过早的被迫选择婚姻,然后,由于盲目的选择,绝大部分的婚姻是不幸的,这也间接的促使了离婚了的增加。另一部分受过高等教育的孩子,会选择晚婚晚育,他们对自身价值的实现过于注重。无论是哪种情况,都不利于我国人口的健康发展。我们应该让正确的婚恋观走进课堂与家庭。加大对正确婚恋观的普及和宣传。

参考文献

- [1] 曹雪. 城市定居对农民工二胎生育意愿的研究[D]: [硕士学位论文]. 南京: 南京林业大学, 2020. <https://doi.org/10.27242/d.cnki.gnjlu.2020.000032>
- [2] 贺丹. 加强战略研究 迎接新时代人口发展挑战[J]. 人口研究, 2018, 42(2): 3-6.
- [3] 杨柠聪. 全面二孩背景下人口生育意愿影响因素研究综述[J]. 重庆社会科学, 2020(1): 94-105.
- [4] 黎友煊. 提升三胎政策效用的企业社会责任[J]. 科技与金融, 2021(7): 67-68.
- [5] 梁同贵. 人口流动会降低生育水平吗?——基于农业户籍流动人口的再检验[J]. 湖南农业大学学报(社会科学版), 2021, 22(6): 29-36. [https://doi.org/10.13331/j.cnki.jhau\(ss\).2021.06.004](https://doi.org/10.13331/j.cnki.jhau(ss).2021.06.004)
- [6] 张孝栋, 张雅璐, 贾国平, 汤梦君, 陈功, 张蕾. 中国低生育率研究进展: 一个文献综述[J]. 人口与发展, 2021, 27(6): 9-21.
- [7] 穆光宗, 林进龙, 江砥. 当代中国人口婚姻嬗变及风险治理[J]. 杭州师范大学学报(社会科学版), 2021, 43(5): 89-97. <https://doi.org/10.19925/j.cnki.issn.1674-2338.2021.05.010>
- [8] 曹园, 方雨, 肖淦丹, 邢佳宁. 养老保险制度、家庭生育决策与社会福利[J/OL]. 南方金融: 1-15. <http://kns.cnki.net/kcms/detail/44.1479.F.20211203.1141.002.html>, 2021-12-07.
- [9] 李飏, 赖德胜, 高曼. 互联网使用对生育率的影响研究[J]. 南方人口, 2021, 36(2): 65-80.
- [10] 倪云松. 房价、房产与生育行为——基于 CHFS 数据的分析[J]. 经济问题, 2021(11): 121-129. <https://doi.org/10.16011/j.cnki.jjw.2021.11.015>
- [11] 何林浩, 陈梦. 夫妻博弈与家庭生育率——对我国生育率下降的一个解释[J]. 世界经济文汇, 2021(4): 74-88.
- [12] 永英. 保障就业合法权益 提升妇女生育意愿[J]. 群众, 2021(14): 31-32.
- [13] 杨成钢, 孙晓海. 中国生育率持续低迷的风险、适应与政策选择[J]. 人口与发展, 2020, 26(4): 101-106+25.
- [14] 毕灵斐. 妇女生育配套福利保障对提高生育率的正向影响探究[J]. 劳动保障世界, 2020(12): 22.
- [15] 张豫南. 家庭功能变迁对女性初婚年龄影响分析——基于 CGSS2015 数据的研究[J]. 河南司法警官职业学院学报, 2020, 18(1): 121-128.