

# 基于变量选择方法的人口出生率影响因素分析

何娟娟

广西师范大学数学与统计学院, 广西 桂林

收稿日期: 2023年12月23日; 录用日期: 2024年2月22日; 发布日期: 2024年2月28日

## 摘要

为响应“三孩”政策, 研究主要影响人口出生率的因素, 并根据各个因素剖析如何促进人口出生率。现有的文献大多是从不同的角度去阐述各个因素分别对人口出生率的影响, 或注重某类因素对人口出生率的影响, 本文在参考已有研究的基础之上, 从经济发展水平、居民生活水平、文化教育和社会保障等方面同时考虑多种因素对人口出生率的影响, 运用变量选择的方法进行实证分析, 并根据分析结果为促进人口出生率提供针对性的建议。

## 关键词

人口出生率, 变量选择方法, 影响因素

## Analysis of Influencing Factors of Population Birth Rate Based on Variable Selection Method

Juanjuan He

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

Received: Dec. 23<sup>rd</sup>, 2023; accepted: Feb. 22<sup>nd</sup>, 2024; published: Feb. 28<sup>th</sup>, 2024

## Abstract

In response to the “three-child” policy, the main factors affecting the birth rate are studied, and how to promote the birth rate is analyzed according to each factor. Most of the existing literatures explain the impact of various factors on the birth rate from different perspectives, or pay attention to the impact of certain factors on the birth rate. Based on the reference of previous studies, this paper simultaneously considers the impact of various factors on the birth rate from the aspects of economic development level, residents' living standard, culture and education, social security, etc.

The method of variable selection is used to make an empirical analysis, and according to the analysis results, specific suggestions are provided to promote the birth rate.

## Keywords

Birth Rate, Variable Selection Method, Influencing Factors

Copyright © 2024 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

从变量选择的角度来看,有限参数设置下的变量选择问题采用惩罚似然方法的变量选择方法,其中包括完善的方法,如赤池信息准则(AIC)和贝叶斯信息准则(BIC),最近的方法,如岭回归[1],以及最小角选择和收缩算子(LASSO)惩罚[2],从对数似然中减去回归系数绝对总和的倍数( $\lambda$ ),将一些回归系数设置为零。Zou [3]结合了 LASSO 和岭回归的优点,通过两种惩罚项(L1 和 L2 正则化项)实现对模型参数的稀疏性和约束。Fan 和 Li 提出一种非凸罚项,称之为平滑裁剪绝对偏差(SCAD)方法[4],其对应的正则化模型不仅能够同时实现变量选择和参数估计,还具有很好的统计性质,即 Oracle 性。我们考虑各变量之间可能存在共线性或模型的复杂性过高,从而考虑利用变量选择中正则化惩罚的方法。一方面正则化方法可以通过约束参数的大小来减少模型的复杂性,从而防止过拟合;另一方面,正则化方法将某些特征的系数压缩为零或很小的值,实现自动特征选择,从而提高预测的准确性。

从已有研究人口出生率的角度看,人口出生率的影响因素错综复杂,不同学者对此研究的切入变量不一样,结果也存在一定的差异。如李勇[5]通过家庭育儿的经济成本、优化医疗资源配置等方面的措施来提升江苏省生育水平,促进人口长期均衡发展。姚文婧[6]测算出各省份数字经济发展水平,分别对中国数字经济与人口出生率现状及地区差异进行理论分析。雷明全[7]分析了婴儿持续减少的直接原因是人们结婚年龄在逐渐延迟和适婚年龄段人口结婚率快速下降。总的来说,现有的文献很少涉及从经济发展水平、居民生活水平、文化教育和社会保障方面同时对人口出生率的影响,本文就这一部分对已有的研究进行补充,力求为提高人口出生率献策。

## 2. 人口出生率的影响因素理论分析

社会层面中教育水平和医疗卫生水平、老年抚养负担和这几个因素在塑造个体和社会整体生育行为方面发挥着关键作用。首先,教育水平在塑造人口出生率中发挥着重要作用。受过良好教育的个体更有可能形成对生育的深刻认识,能够理解生育决策的后果。教育的提升往往伴随着更广泛的职业选择和更好的职业机会,因此个体可能更倾向于推迟生育,以追求事业目标。女性在受过良好教育的社会中通常享有更高的地位,她们更可能追求事业和自主权,从而影响生育决策的谨慎性。其次,医疗卫生水平也对人口出生率产生显著影响。在医疗条件良好的地区,人们更愿意冒险生育,因为婴儿在生存方面的机会更高。医疗卫生水平的提高有助于改善婴儿的存活率,这可能影响个体对于生育的态度。此外,良好的医疗卫生水平通常伴随着更好的生育健康,为个体提供了更安全的生育环境。医疗服务的普及也使人们更容易获取避孕信息和工具,从而更灵活地控制生育。第三,老年抚养比(老年人口数量与劳动年龄人口数量的比例)是人口结构的一个重要指标,对人口出生率产生着深远的影响。老年抚养比高时,家庭通常需要承担更多的老年抚养责任,高老年抚养比可能会对人口出生率产生压力,导致个体和家庭更倾向

于控制生育率，以适应老龄化社会的经济和资源压力。

从人口层面来看，适婚的男性与女性的性别比例、人口生育观念等都对人口出生率起着至关重要的作用。首先，性别比例在特定文化和社会背景下对人口出生率产生独特的影响，性别比例失衡可能导致婚姻市场的变化，使得适婚的男性与女性失调的地区出现婚姻匹配的竞争激烈。这种情况下，家庭可能更倾向于控制生育，以适应有限的婚配资源。同时，性别比例的失衡也可能激发对未来性别平衡的担忧，促使一些家庭选择限制生育或采取性别选择的手段，从而影响到整体的出生率。其次，人口生育观念在个体和家庭层面塑造着生育决策。传统的生育观念可能鼓励大家多生多育，将子女视为养老的依靠。然而，在现代社会，随着经济发展和教育水平提高，人们的生育观念逐渐转变。家庭可能更注重子女的教育品质而非数量，导致生育率的降低。女性的地位提高也使得她们更能够参与社会 and 职业生活，可能更愿意推迟生育。性别比例和生育观念的变化都在一定程度上反映着社会的进步和演变，直接或间接地影响到个体对于生育的态度和实际行为。因此，深入理解和分析人口性别比例和生育观念对人口出生率的作用，对于有效制定相关政策和促进社会可持续发展至关重要。

从经济发展层面和政策调控层面来讨论对人口出生率的影响。首先，随着经济水平的提高，家庭的收入水平和生活质量得到改善，人们更有能力和愿望投资于子女的教育和健康。这可能导致生育率的下降，因为家庭更倾向于注重子女少而优的教育质量，而非数量。其次，政策调控层面对人口出生率同样有着重要作用。政府制定的相关政策，如计划生育政策、家庭政策等，直接影响到个体和家庭的生育决策。通过实施鼓励生育的政策、提供生育奖励或津贴等措施，应对人口老龄化和出生率下降的挑战。这些政策可能通过激励或限制的手段影响到家庭的生育意愿和行为，进而对整体出生率产生影响。

### 3. 人口出生率影响因素实证研究

数据来源于国家统计局和各省的统计年鉴，选取 2011 年至 2022 年的数据，将得到的数据进行整合，对于缺失的数据，本文均运用均值插补法进行插补，有效地处理了缺失值异常值等问题。同时，含有地区和年份的数据通过引入虚拟变量的方式，使得在同一模型下会有不同的参数或反映不同省份之间差异的大小，更准确地分析各个变量对响应变量的影响，有助于模型收敛并提高解释性。

#### 3.1. 模型设定

建立多元线性回归模型，

$$Y = X\beta^* + \varepsilon$$

其中， $X = (X_1, X_2, \dots, X_p)^T$ ， $\beta^* = (\beta_1^*, \beta_2^*, \dots, \beta_p^*)^T$ ， $p$  是变量的维度， $\varepsilon$  是服从正态分布的随机误差项。为探究各变量与  $Y$  之间的关系，考虑施加正则化惩罚后需对感兴趣的参数  $\beta$  进行估计，建立加上 SCAD 惩罚的目标函数，

$$f_2(\beta) = (Y - X^T\beta)^T (Y - X^T\beta) + p_\lambda(\beta),$$

这里的  $p_\lambda(\beta)$  是 SCAD 惩罚函数，

$$p_\lambda(\beta) = \begin{cases} \lambda|\beta_j|, & 0 \leq |\beta_j| < \lambda \\ (-1) * \frac{|\beta_j|^2 - 2a\lambda|\beta_j| + \lambda^2}{2(a-1)}, & \lambda \leq |\beta_j| < a\lambda \\ \frac{\lambda^2(a+1)}{2}, & |\beta_j| \geq a\lambda \end{cases}$$

其中，超参数在惩罚函数中发挥关键作用，用于控制模型的惩罚强度。选择适当的参数对模型结果有着重要影响。通常，我们采用 AIC、BIC 以及交叉验证 CV 等三种标准来进行参数调整。为了确定本文中四种变量选择方法的最优正则化参数，我们使用了交叉验证 CV 准则。

### 变量选取

选取人口出生率(Y)作为响应变量，即被解释变量，来衡量我国人口出生情况变动水平。

人口出生率的变化不仅由单一因素所决定，是受到各种不同层面的因素影响，包括经济、社会、人口结构以及政策等。本研究将综合考虑这些因素，并结合广泛的人口相关文献及专家学者的研究分析，选取了一系列可靠的指标和变量进行探讨。具体而言，从经济、社会和人口、环境出发，我们选取了全体居民人均消费支出( $X_1$ )、住宅商品房销售额( $X_2$ )、农村人均可支配收入( $X_3$ )、城镇人均可支配收入( $X_4$ )、城镇化率( $X_5$ )、失业率( $X_6$ )、国内专利申请受理量( $X_7$ )、地方财政房产税( $X_8$ )这几个具有代表意义的变量，用以评估各方面的经济发展水平；粗离婚率( $X_9$ )、结婚人数( $X_{10}$ )这三个相关变量来衡量与出生率相关的人口结构水平；用 15 岁及以上文盲男性人数比上女性人数(抽样调查)( $X_{11}$ )表示社会的受教育程度；用年末参加生育保险的人数( $X_{12}$ )这个指标描述对生育的社会保障；选取治理废水项目完成投资( $X_{13}$ )，治理废气项目完成投资( $X_{14}$ )衡量环境层面对人口出生率的影响；社会发展层面用医疗卫生机构床位数( $X_{15}$ )，每万人拥有卫生技术人员数( $X_{16}$ )、老年人口抚养比( $X_{17}$ )。综上，选取 17 个变量(X)作为影响人口出生率的因素，即是  $X_1, X_2, \dots, X_{17}$  表示解释变量或称为协变量。

### 3.2. 实证结果及分析

**Table 1.** The result of parameter estimation

**表 1.** SCAD 参数估计结果

Intercept	10.4682527
X1	-4.619536
X2	0.00
X3	0.00
X4	3.9619228
X5	-1.0825547
X6	-0.1433638
X7	0.00
X8	0.00
X9	0.00
X10	0.00
X11	-0.5826107
X12	0.00
X13	0.00
X14	0.00
X15	0.00
X16	0.00
X17	-1.3153889

由结果分析表 1 可看出，SCAD 惩罚后的参数估计结果可以看出全体居民人均消费支出、城镇居民人均可支配收入、城镇率、失业率、15 岁及以上文盲男性人数比上女性人数(抽样调查)、老年人口抚养比这几个指标对人口出生率有较为明显的影响。其中，城镇居民人均可支配收入的增加会促进出生率。

城镇率、失业率、粗离婚率、15岁及以上文盲男性人数比女性人数，以及老年抚养比的增加时会使得人口出生率的减少。在研究居民人均消费支出与人口出生率的关系时，我们注意到物价的波动可能直接影响家庭的生活水平。高层次的消费价格通常伴随着更高的生活成本，可能使家庭更加谨慎地考虑生育，随着居民消费价格指数的上升，人口出生率可能会减少。经济条件较好的家庭更有能力承担生育和抚养子女的责任，城镇人均可支配收入的增加可能刺激人们对生育的积极性，从而导致人口出生率的增加；城镇率和失业率的上升可能导致家庭经济能力降低，从而影响生育的意愿；粗离婚率的增加可能表明家庭结构的不稳定，进而影响到生育决策；老年人口的增加，家庭可能需要更多的资源来满足养老需求，家庭经济压力增加，使得年轻夫妇推迟生育或者减少生育意愿。

## 4. 结论和建议

### 4.1. 结论

全体居民人均消费支出的增加会带动人口出生率的增加。城镇率、失业率、粗离婚率、15岁及以上文盲男性人数比女性人数、老年人口抚养比的增加时会使得人口出生率的减少。从这几个角度出发给出解决人口出生率下降建议，但是人口出生率下降的问题除了需要综合考虑经济、教育、社会和文化等多个方面的因素，还需要政府、社会各界和个人共同努力，共同创造一个支持生育的环境，以实现人口出生率的积极增长，为社会的持续健康发展提供基础和保障。

### 4.2. 提出建议

经济政策调整。通过实施有利于家庭经济状况的政策，例如减税、提高家庭津贴或其他经济激励措施，以鼓励年轻夫妇考虑生育。确保居民消费价格指数的稳定，防止通货膨胀对家庭经济的负面影响提高教育水平：投资提高全民教育水平，特别是女性教育。教育的提升通常与延迟生育决策、更理性的生育观念以及更好的家庭规划有关。通过降低文盲率和提高女性的受教育水平，有望推动人口出生率的增加。

家庭支持与照顾服务。可提供全面的家庭支持和照顾服务，包括弹性的工作时间、产假政策和优质的儿童托管服务。这些措施有助于缓解年轻父母的生活压力，使他们更愿意考虑生育。

性别平等政策。推动性别平等，包括在职场和家庭中平等对待男女。减少失业率和城镇率对生育决策的负面影响，确保女性有更多的职业选择和机会。

加强社会支持网络。建设更完善的社会支持网络，包括心理健康服务、家庭辅导和社区互助。这有助于应对城市化和家庭结构变化对人们生育决策的影响，提高生育的社会支持。

文化宣传与教育。进行有针对性的文化宣传和教育活动，强调家庭的重要性，并促使社会更加积极地看待生育。通过塑造积极的生育价值观，可以改变人们对生育的态度，从而提高人口出生率。

## 参考文献

- [1] Hoerl, A.E. and Kennard, R.W. (2000) Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, **42**, 80-86. <https://doi.org/10.1080/00401706.2000.10485983>
- [2] Tibshirani, R. (1996) Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society: Series B*, **58**, 267-288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- [3] Zou, H. and Hastie, T. (2005) Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **67**, 301-320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>
- [4] Fan, J. and Li, R. (2001) Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties. *Journal of the American Statistical Association*, **96**, 1348-1360. <https://doi.org/10.1198/016214501753382273>

- [5] 李勇, 孙琪. 如何有效促进人口生育水平提升——基于江苏省人口出生率影响因素的实证分析[J]. 中国卫生事业管理, 2023, 40(7): 486-490.
- [6] 姚文婧. 数字经济对人口出生率的影响研究[D]: [硕士学位论文]. 沈阳: 辽宁大学, 2023.
- [7] 雷明全, 胡武贤. 我国人口出生率下降的教育困境[J]. 长春大学学报, 2023, 33(4): 44-49+54.