

# 我国人口生育行为影响因素的多模型分析研究

许鑫颖, 尹勤\*, 张菊, 马嘉琪

南京邮电大学经济学院, 江苏 南京

收稿日期: 2023年6月18日; 录用日期: 2023年7月13日; 发布日期: 2023年7月24日

## 摘要

长期极低的生育率将会造成人口失衡和劳动力短缺的后果。本文基于中国社会状况综合调查(CSS)数据, 运用有序Logistic回归、Poisson回归、LightGBM模型、中介效应模型等方法, 对不同特征育龄群体的生育行为进行差异分析并探究影响生育行为的社会经济因素, 主要研究结论有: 1) 当前有约四成的育龄群体选择终身生育1个或不生孩子, 除性别外, 不同特征育龄群体之间生育行为差异显著。2) 受教育水平越高、家庭人均收入越高、家庭相对收入越低、家庭规模越小、参与养老保险的育龄群体其生育更多子女的可能性更低, 其中, 家庭规模、个人及家庭收入的特征重要性更高。3) 中介效应结果显示, 受教育程度、家庭规模会通过影响生育意愿, 从而对生育行为产生影响。为此本文提出了相应建议, 为改善生育水平, 促进人口均衡发展提供决策参考。

## 关键词

生育行为, 社会经济因素, 特征重要性, 中介效应

# Multi-Model Analysis and Research on Influencing Factors of Population Fertility Behavior in China

Xinying Xu, Qin Yin\*, Ju Zhang, Jiaqi Ma

School of Economics, Nanjing University of Posts and Telecommunications, Nanjing Jiangsu

Received: Jun. 18<sup>th</sup>, 2023; accepted: Jul. 13<sup>th</sup>, 2023; published: Jul. 24<sup>th</sup>, 2023

## Abstract

Chronically very low fertility rates will lead to population imbalances and labor shortages. Based on the data of Chinese Social Survey (CSS) and using the methods of ordered Logistic regression,

\*通讯作者。

文章引用: 许鑫颖, 尹勤, 张菊, 马嘉琪. 我国人口生育行为影响因素的多模型分析研究[J]. 应用数学进展, 2023, 12(7): 3277-3291. DOI: 10.12677/aam.2023.127327

Poisson regression, LightGBM model and intermediary effect model, this paper analyzes the differences in fertility behaviors of different characteristics of childbearing age groups and explores the socio-economic factors affecting fertility behaviors. The main conclusions are as follows: 1) Currently, about 40 percent of people of childbearing age choose to have one or no children at all. In addition to gender, there are significant differences in fertility behavior among different characteristics of childbearing age groups. 2) The higher the level of education, the higher the per capita income of the family, the lower the relative income of the family, the smaller the family size, and the lower the probability of having more children in the childbearing age group participating in the pension insurance. Among them, the characteristics of family size and family income are more important. 3) The mediating effect model shows that education level and family size affect reproductive behavior through influencing reproductive intention. Therefore, this paper puts forward corresponding suggestions to provide decision-making reference for improving fertility level and promoting balanced population development.

## Keywords

Fertility Behavior, Socio-Economic Factors, Feature Importance, Mediating Effect

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

国家统计局最新资料显示, 2022 年我国 60 岁及以上的老年人口数为 2.80 亿, 占总人口的 19.8% [1], 老龄化程度不断加深。同时, 近年来我国出生人口和生育水平呈现持续低迷的态势。2022 年我国人口进入生育主导的人口负增长阶段, 人口出生率降至 6.77‰, 且据第七次人口普查数据, 2020 年我国出生人口减少至 1200 万人, 总和生育率仅为 1.30, 远低于生育更替水平 2.1 [2]。由此可见, 当前我国老龄少子化问题凸显。我国政府也在不断优化人口生育政策以平衡人口年龄结构, 从“单独二孩”到“全面二孩”政策以及 2021 年全国人大常委会提倡的“一对夫妻生育三个子女”的政策, 均旨在改善生育水平, 促进人口均衡发展。长期极低的生育率将会造成人口失衡和劳动力短缺的后果, 因此, 有必要对我国育龄群体生育行为进行全面了解, 并进一步探究影响生育行为的相关因素。

## 2. 文献综述

生育行为是育龄群体所产生的实际生育行为, 是人们考量生育意愿和生育成本后, 所进行的生育选择[3]。生育行为可以拆分成数量、性别、时间等维度, 具体包括子女数量、性别结构、生育时间、胎次间隔等要素[4]。生育行为的相关研究主要围绕其影响因素展开, 刘璐婵和张瑞利的研究结果显示, 教育成本对生育行为有抑制效应, 且农村、低收入和东部地区对于教育成本更为敏感。中介效应结果显示教育开支能降低父母的预期子女数, 从而扭转其生育意愿, 导致消极的生育行为[5]。冯永琦和于欣晔的研究认为, 家庭投资风险偏好会对生育行为产生影响, 较高的投资风险偏好会对子女的生育行为产生抑制作用。此外, 收入水平、受教育程度、户籍状况、年龄、拥有房产数量以及收入风险等因素对生育行为也有较大的影响[6]。顾和军等运用 CHNS 调查资料, 针对祖辈“退而不休”现象对育龄妇女生育行为的影响及其机理进行了研究, 结果显示, 在其它条件不变的前提下, 祖辈“退而不休”明显降低了育龄妇女的生育数量, 在对该模型处理内生性后, 回归分析结果仍然显著[7]。现有的研究中, 生育行为的影响因素研究多局限在验证一个或两个因素对其的影响, 缺少更为全面的社会经济影响因素探索分析, 本文

将结合传统的计量经济学模型、机器学习分类算法和中介效应模型,从个体、家庭、社会保障、地区特征四个方面探究生育行为的影响因素,不但充实了现有的研究方法体系,也更为全面地探究了影响生育行为的因素。

### 3. 数据来源与模型构建

#### 3.1. 数据来源

本文的数据来源为中国社会状况综合调查(CSS)数据,中国社会状况综合调查采用入户抽样的访问方式,调查区域覆盖了除港澳台、新疆外的全国 30 个省(自治区、直辖市),包括了 151 个区市县,604 个村/居委会,每项调查访问了 7000 至 10,000 多个家庭,以 18~69 周岁以上住户人口为调查对象。CSS 调查问卷从设计上分三大部分,即基础模块,更替模块,热点模块。其中基础模块是固定的,包括个人的基础信息,工作和就业,家庭结构和家庭经济状况等;更替模块比如社会阶层地位的流动,社会保障,休闲消费和社会价值观,在间隔一定周期后进行重复调查;热点模块紧跟时代步伐,话题涉及但不局限于社会群体利益关系,民生问题和城镇化问题等。本文选取 2019、2021 两年的数据进行了合并与汇总,共获得有效样本 3659 个。

#### 3.2. 模型选择

##### 3.2.1. 有序 Logistic 回归

有序 logistic 回归,是根据有序多分类变量拆分成多个二分类因变量,拟合多个二分类 logistic 回归,并基于累积概率构建回归模型。本文的因变量为生育行为(生育 0 或 1 个、生育 2 个、生育 3 个及以上),则因变量可以拆分成两个二分类回归模型,一个是生育 0 或 1 个与生育 2 个及以上,另一个则为生育 2 个及以下和生育 3 个及以上。计算公式如式(1)所示。

$$\text{logit}[P(Y \leq i | X)] = \ln \left[ \frac{P(Y \leq i | X)}{1 - P(Y \leq i | X)} \right] = \beta_{0i} + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m \quad (1)$$

式中,  $i$  代表因变量的第  $i$  个类别,  $i = 1, 2, 3$ ;  $Y$  代表因变量生育行为,  $X$  代表自变量;  $\beta_{0i}$  为常数项系数,  $\beta_1, \beta_2, \dots, \beta_m$  分别为自变量系数。

有序 logistic 满足比例优势假设,在拆分的多个二分类 logistic 回归中,假定除了截距不同,自变量对应的模型系数均相等,故累计概率的优势比(OR)值也相等。

##### 3.2.2. Poisson 回归

当因变量为计数型变量时,使用 Poisson 回归或者负二项回归,而负二项回归模型使用的前提条件是因变量的方差大于期望,本文生育行为的方差小于期望,依据最大似然比检验判定选用 Poisson 模型[8]。Poisson 回归模型的具体设定分别如下:

$$\Pr(\text{fxingwei} = n_i | x_i) = \frac{e^{-\lambda_i} \lambda_i^{n_i}}{n_i!} \quad (n_i = 0, 1, 2, \dots, 12) \quad (2)$$

$$\lambda_i = E(\text{fxingwei} = n_i | x_i), \quad x_i = \exp(X_i' \beta) \quad (3)$$

##### 3.2.3. LightGBM 分类算法

LightGBM 模型一种基于梯度提升树算法的改进模型,具有精度高、可处理海量数据等优点。该模型主要是利用决策树迭代训练以获得最优模型,计算见式(4)。

$$F_g(X) = \sum_{k=1}^g f_k(X) \quad (4)$$

式中： $F_g(X)$ 为 $g$ 棵决策树形成的模型； $X$ 为自变量， $f_k(X)$ 为第 $k$ 棵决策树。

该模型的最终目的是要让识别结果更接近于真实结果，因此，在每次迭代的过程中，都要找到决策树，以使得目标函数的值达到最小。其中，目标函数包含了两个部分，一个是损失函数，另一个是正则化项。损失函数的计算见式(5)。

$$L(y_q, y'_q) = \frac{1}{A} \sum_{q=1}^A y_q \lg P_q + (1 - y_q) \lg (1 - P_q) \tag{5}$$

式中： $L(y_q, y'_q)$ 为损失函数； $y_q$ 为自变量 $X_q$ 对应的真实类别； $y'_q$ 为输入自变量 $X_q$ 后识别的类别； $A$ 为输入样本量； $P_q$ 为将自变量 $X_q$ 输入模型后识别为类别1的概率。

正则化项通过控制叶子节点分裂以降低模型的过拟合，该模型的目标函数见式(6)。

$$O = L(y_q, y'_q) + \gamma Z + \frac{1}{2} \lambda \sum_{v=1}^Z (W_v)^2 \tag{6}$$

式中： $O$ 为目标函数； $Z$ 为叶子节点数； $W_v$ 为第 $v$ 个叶子节点的输出值； $\gamma$ 和 $\lambda$ 是设定的参数。

相比较传统的梯度提升算法，LightGBM模型改进为采用直方图的决策树算法，并与受深度限制的叶子生长策略相结合，从而获得了更高的精度和运算速度。且该模型使用了单边梯度采样算法，通过增加大梯度样本的权重，降低小梯度样本的权重来计算加权信息增益[9]。

### 3.2.4. 中介效应模型

中介效应模型是检验某一变量是否成为中介变量，发挥何种程度中介作用的一种方法。变量之间的关系如下，图1(a)是自变量 $X$ 作用于因变量 $Y$ ，路径系数为 $c$ ，在没有中介变量的情况下，系数 $c$ 反映了自变量对因变量的影响程度。图1(b)显示了对中间变量 $M$ 进行调节之后，自变量 $X$ 与因变量 $Y$ 的关系，系数 $a$ 指自变量对中间变量的影响， $b$ 指中间变量对因变量的影响。系数 $c'$ 是指对中间变量进行控制后，自变量对因变量产生的影响。图1(b)中的变量间总效应等于直接效应加上间接效应，即 $c = ab + c'$ ， $c$ 为总效应， $c'$ 为直接效应， $ab$ 为中介效应，也可称为间接效应。

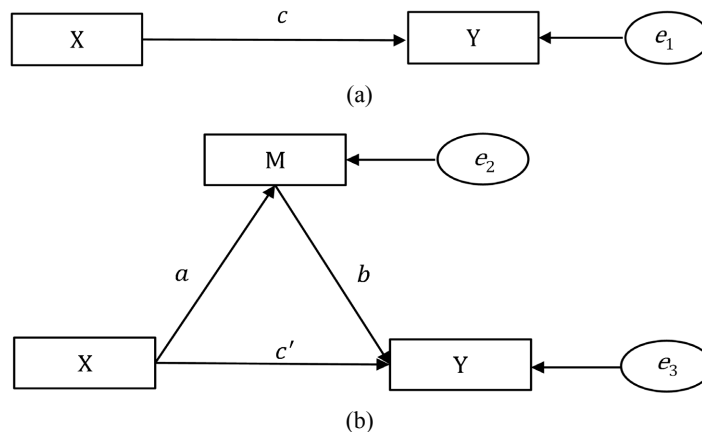


Figure 1. The diagram of test path of mediation effect  
图 1. 中介效应的检验路径示意图

## 3.3. 变量设置

### 3.3.1. 变量说明

本文取 40~49 岁育龄群体作为研究对象，该年龄段的生育行为尚未完全结束，但依据第七次全国人

口普查数据, 40~49 岁育龄群体年龄别生育率仅为 7.95%, 且结合实际, 这一年龄段的育龄人群大多数都终止了生育行为, 故本文聚焦 40~49 岁育龄人群, 将这一群体的实际生育孩子数作为当前育龄人群终身生育子女数的近似估计。

1) 因变量。本文的因变量为生育行为, 数据来源于问卷中“您现在有几个亲生子女?”的问题。将生育数量为 0 或 1 个定义为 1, 2 个定义为 2, 大于等于 3 个定义为 3。

2) 中介变量。本文的中介变量为生育意愿, 数据来源于问卷中“您认为一个家庭通常有几个孩子最理想?”的问题, 也即育龄群体的理想子女数。

3) 自变量。本文的自变量主要选取个体属性、家庭属性、社会保障、地区属性四个方面, 将受教育程度、家庭规模、家庭人均收入、家庭相对收入、养老保险、医疗保险作为主要的社会经济解释变量, 其中, 家庭规模数据来源于“请您告诉我您家有几口人?”这一问题, 家庭人均收入由家庭总收入除以家庭规模计算得到(家庭总收入来源于“您家的总收入”这一问题), 家庭相对收入[7]为家庭人均收入除以所在区/县各家庭人均收入的中位数, 家庭人均收入反映的是家庭的绝对收入水平, 而家庭相对收入反映的是家庭的相对收入水平。相关变量描述性统计如表 1 所示。

**Table 1.** Description of variables

**表 1.** 变量说明

变量类型		变量名称	变量赋值	均值	标准差
因变量	——	生育行为	1 = 0 或 1 个, 2 = 2 个, 3 = 大于等于 3 个	1.71	0.67
中介变量	——	生育意愿	数值型变量	3.09	10.11
自变量	个体属性	性别	1 = 男, 2 = 女	1.59	0.49
		年龄	数值型变量	44.87	2.88
		受教育程度	1 = 小学及以下, 2 = 初中, 3 = 高中/中专/职高, 4 = 本科/大专, 5 = 研究生	2.20	1.06
		婚姻状况	1 = 初婚有配偶, 2 = 再婚有配偶, 3 = 离婚, 4 = 丧偶	2.20	0.58
		个人收入	数值型变量	29506.55	27251.00
	家庭属性	家庭规模	数值型变量	4.34	1.63
		家庭相对收入	数值型变量	1.33	1.17
		家庭人均收入	数值型变量	17075.25	14451.27
	社会保障	养老保险	0 = 否, 1 = 是	0.45	0.50
		医疗保险	0 = 否, 1 = 是	0.76	0.42
	地区属性	所在地区	1 = 华北, 2 = 东北, 3 = 华东, 4 = 中南, 5 = 西南, 6 = 西北	3.50	1.36
		城乡居住属性	1 = 城市, 2 = 乡村	1.40	0.49

### 3.3.2. 异常值处理

选择 KS 检验对表 1 中各数值型变量进行正态性检验, 检验结果如表 2 所示, 可以看到, 各变量的 P 值均小于显著性水平 0.05, 拒绝“服从正态分布”的原假设, 故认为各数值型变量均不服从正态分布。

**Table 2.** Normality test of numerical variables**表 2.** 数值型变量的正态性检验

变量	KS 统计量	P 值
生育意愿	0.113	0.000
年龄	0.088	0.000
家庭规模	0.199	0.000
个人收入	0.139	0.000
家庭相对收入	0.160	0.000
家庭人均收入	0.137	0.000

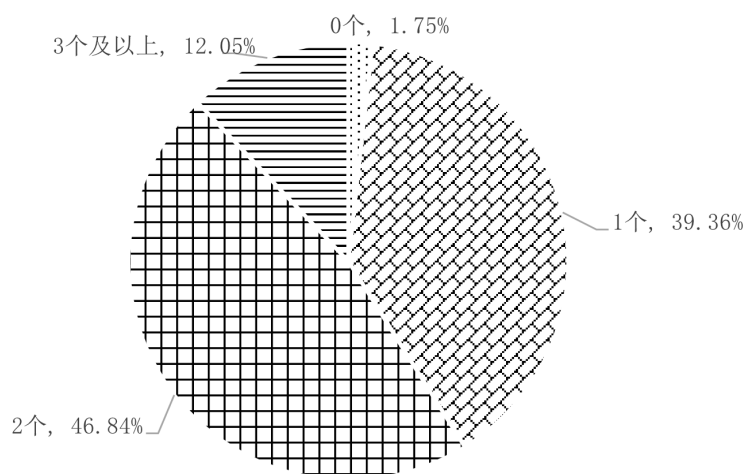
因此本文选取箱线法对所有数值型的变量进行异常值检测，若直接对异常值对应的样本进行删除处理，则本文保留的数据量会大量缩减，为保留其他变量中所包含的有效信息，故本文采取将异常值视为缺失值的方法来进行异常值的处理。

### 3.3.3. 空值处理

本文对分类型变量，采用众数填补缺失值的方法，对于数值型变量，采用中位数填补的方法。

## 4. 不同特征人群生育行为差异分析

育龄妇女 2019、2021 年平均终身生育子女数分别为 1.71、1.74，与对应年份理想子女数均相差 0.29，可见生育意愿与生育实际之间存在一定的差异。图 2 是调查对象的终身生育子女数量分布，可以看出，当前育龄群体没有生育孩子的样本量占比 1.75%，生育 1 孩的占比 39.36%，选择终身生育 2 个孩子的样本量占比为 46.84%，选择生育 3 个及以上的孩子占比为 12.05%。

**Figure 2.** Distribution of lifetime births**图 2.** 终身生育子女数分布

### 4.1. 个体特征间的差异分析

表 3 展示了不同个体特征之间生育行为差异分布及卡方检验结果。分性别来看，不同性别育龄群体拥有的子女数无明显差异。分年龄来看，40~44 岁育龄人群拥有 2 个、3 个及以上的样本占比分别为 48.98%、13.08%，分别高出 45~49 岁育龄人群 3.82、1.84 个百分点，两个年龄段之间的生育行为存在显

著差异。就受教育程度来看,小学及以下育龄群体生育 2 孩及以上的占比较高,为 77.26%,受教育程度为本科/大专、研究生的育龄群体生育 2 孩及以上占比较少,分别为 34.68%、35.29%,不同受教育程度育龄群体之间的生育行为差异显著。婚姻状况为已婚的育龄群体生育 2 个及以上孩子的占比较高,离婚/丧偶的育龄群体则较低,不同婚姻状况育龄群体之间的生育行为差异显著。就个人收入而言,收入较低的育龄群体生育 2 个及以上孩子的样本占比较高,收入中等的占比较低,而收入较高的育龄群体生育 2 个及以上孩子的样本占比反而高于收入中等的群体。由此推测,当个人收入高于一定水平时,育龄人群的生育行为将会更多。

**Table 3.** Distribution of reproductive behaviors with different individual characteristics  
**表 3.** 不同个体特征的生育行为分布

个体特征	变量分组	0 个或 1 个		2 个		3 个及以上		卡方值
		频数(人)	占比(%)	频数(人)	占比(%)	频数(人)	占比(%)	
性别	男	629	41.91	704	46.90	168	11.19	1.961
	女	875	40.55	1010	46.80	273	12.65	
年龄	40~44 岁	612	37.94	790	48.98	211	13.08	12.355***
	45~49 岁	892	43.60	924	45.16	230	11.24	
受教育程度	小学及以下	249	22.74	607	55.43	239	21.83	397.228***
	初中	530	39.64	664	49.66	143	10.70	
	高中/中专/职高	349	53.61	259	39.78	43	6.61	
	本科/大专	354	65.31	172	31.73	16	2.95	
	研究生	22	64.71	12	35.29	0	0.00	
婚姻状况	初婚有配偶	1265	39.46	1549	48.32	392	12.23	98.149***
	再婚有配偶	77	35.48	113	52.07	27	12.44	
	离婚	142	73.58	39	20.21	12	6.22	
	丧偶	20	46.51	13	30.23	10	23.26	
个人收入	<1 w	277	28.62	531	54.86	160	16.53	127.320***
	1~3 w	441	40.87	495	45.88	143	13.25	
	3~5 w	380	46.57	349	42.77	87	10.66	
	5~7 w	207	48.71	183	43.06	35	8.24	
	7~9 w	107	57.84	71	38.38	7	3.78	
	≥9 w	92	49.46	85	45.70	9	4.84	

注:\*\*\*、\*\*、\*分别代表 1%、5%、10%的显著性水平,下同。

#### 4.2. 家庭特征间的差异分析

不同家庭特征育龄群体之间的生育行为存在显著差异。如表 4 所示,家庭规模为 4 个及以上的育龄群体生育 2 个或以上的占比较高。将相对收入小于 0.5 归类为“很低”,大于等于 0.5 且小于 0.8 归类为“较低”,大于等于 0.8 且小于 1.2 归类为“中等”,大于等于 1.2 且小于 1.5 归类为“较高”,大于等于 1.5 归类为“很高”,可得相对收入与生育行为呈“U”型的数量关系,相对收入很高和相对收入很低的育龄人群生育 2 孩及以上的样本占比均较高,因此家庭相对收入不同,生育行为也有所不同。家庭人均收入越高,生育 2 个及以上孩子的样本占比就越低。

**Table 4.** Distribution of reproductive behavior of different family characteristics  
**表 4.** 不同家庭特征的生育行为分布

家庭特征	变量分组	0 个或 1 个		2 个		3 个及以上		卡方值
		频数(人)	占比(%)	频数(人)	占比(%)	频数(人)	占比(%)	
家庭规模	1~2 人	129	100.00	0	0.00	0	0.00	2200.000***
	3 人	861	89.78	98	10.22	0	0.00	
	4 人	224	19.55	896	78.18	26	2.27	
	5 人	171	24.89	338	49.20	178	25.91	
	6 人	48	12.44	248	64.25	90	23.32	
	7 人及以上	48	17.39	108	39.13	120	43.48	
家庭相对收入	低	212	30.72	341	49.42	137	19.86	100.543***
	较低	248	38.27	308	47.53	92	14.20	
	平均水平	551	50.41	469	42.91	73	6.68	
	较高	205	44.96	207	45.39	44	9.65	
	高	288	37.31	389	50.39	95	12.31	
家庭人均收入	<1 w	305	25.17	652	53.80	255	21.04	401.544***
	1~2 w	494	38.09	676	52.12	127	9.79	
	2~3 w	334	56.13	223	37.48	38	6.39	
	≥3 w	371	66.85	163	29.37	21	3.78	

### 4.3. 社会保障特征间的差异分析

表 5 结果显示, 参与/未参与养老保险育龄群体之间的生育行为存在显著差异。相较于未参与, 参与养老保险的育龄群体生育 2 个、3 个及以上孩子的占比较少, 分别为 41.85%、8.30%, 较未参与的低了 9.15、6.88 个百分点。参与医疗保险的育龄群体生育 0 个或 1 个的育龄群体占比 42.35%, 相较于未参与养老保险高出 6.22 个百分点, 参与/未参与医疗保险育龄群体之间的生育行为存在显著差异。

**Table 5.** Distribution of reproductive behaviors with different social security characteristics  
**表 5.** 不同社会保障特征的生育行为分布

社会保障特征	变量分组	0 个或 1 个		2 个		3 个及以上		卡方值
		频数(人)	占比(%)	频数(人)	占比(%)	频数(人)	占比(%)	
养老保险	否	675	33.82	1018	51.00	303	15.18	108.589***
	是	829	49.85	696	41.85	138	8.30	
医疗保险	否	313	36.35	450	52.26	98	11.38	13.668***
	是	1191	42.57	1264	45.18	343	12.26	

### 4.4. 地区特征间的差异分析

不同地区育龄人群的生育行为差异显著。表 6 结果显示, 居住在农村地区的育龄人群生育 2 个及以上的占比 72.48%, 高于城市 22.57 个百分点。所在地区为中南、西南、西北的育龄人群生育 2 孩、3 孩及以上的占比均较高, 说明这些地区生育多孩的现象较为普遍。



**Table 6.** Distribution of reproductive behavior in different regions  
**表 6.** 不同地区特征的生育行为分布

地区特征	变量分组	0 个或 1 个		2 个		3 个及以上		卡方值
		频数(人)	占比(%)	频数(人)	占比(%)	频数(人)	占比(%)	
城乡	城市	1103	50.09	929	42.19	170	7.72	220.339***
	乡村	401	27.52	785	53.88	271	18.60	
所在地区	华北	210	49.41	194	45.65	21	4.94	287.078***
	东北	211	73.01	71	24.57	7	2.42	
	华东	502	46.92	477	44.58	91	8.50	
	中南	295	28.67	531	51.60	203	19.73	
	西南	197	33.45	301	51.10	91	15.45	
	西北	89	34.63	140	54.47	28	10.89	

## 5. 育龄人群生育行为影响因素研究

本文从个体特征、家庭特征、社会保障、地区因素四个方面探究影响育龄群体生育行为的因素。重点研究受教育程度、家庭支持(家庭人均收入、家庭相对收入、家庭规模)、社会保障(是否参与养老保险、是否参与医疗保险)等社会经济因素对生育行为的影响路径和大小,上述变量为主要的解释变量,其余为控制变量。

首先使用方差膨胀因子(VIF)对自变量之间的自相关性进行了检验,各解释变量的方差膨胀因子如表 7 所示,根据方差膨胀因子判断准则( $VIF < 5$  时自变量之间不存在自相关),所有自变量中最大的方差膨胀因子也仅为 2.68,故各变量之间不存在自相关性。

**Table 7.** Variance inflation factors  
**表 7.** 方差膨胀因子

自变量	VIF
家庭人均收入	2.68
家庭相对收入	2.20
受教育程度	1.51
养老保险	1.31
城乡	1.28
家庭规模	1.26
个人收入	1.17
调查年份	1.16
医疗保险	1.16
性别	1.11
所在地区	1.09
年龄	1.05
婚姻状况	1.04
Mean VIF	1.39

## 5.1. 基准回归结果与稳健性检验

### 5.1.1. 基准回归结果

表 8 展示的是有序 logistic 回归的结果，主要解释变量中，受教育程度、家庭人均收入、家庭规模、家庭相对收入、养老保险对被解释变量生育行为影响显著，医疗保险变量对生育行为不显著；控制变量中年龄、婚姻状况、个人收入、城乡、所在地区、年份对生育行为影响显著，而性别对生育行为影响不显著。

相较于小学及以下，受教育水平高的育龄群体，其选择多生育孩子的可能性较低，学历为初中的育龄群体多生育孩子的可能性相较于小学及以下降低了 44.9%，同理，高中/中专/职高、本科/大专、研究生分别下降了 61.0%、74.4%、66.8%，对比得到，学历为本科/大专的育龄群体产生更多生育行为的可能性最低，研究生学历的产生更多生育行为的可能性较本科/大专的育龄群体高。

家庭支持包含了人力、财力两个方面的变量。其中，家庭规模越大，获得的家庭生育支持越多，产生更多生育行为的可能性就越大。家庭人均收入越高，产生更多生育行为的可能性越小。而家庭的相对收入越高，产生更多生育行为的可能性则越大。

社会保障变量中，参加养老保险会降低更多生育行为发生的可能性，而参加医疗保险对生育行为的影响不显著。

控制变量中，年龄越大，生育更多孩子的可能性越小，推测年龄越大，育龄人群处在计划生育阶段的时期就越长。婚姻状况为离婚的育龄群体相较于在婚人群生育更多孩子的可能性较低，而丧偶的育龄群体生育更多孩子的可能性更高。个人收入对生育行为的影响显著，个人收入越高，产生更多生育行为的可能性就越低。居住在农村地区的育龄妇女相较于城市地区生育更多子女的可能性更高。相较于华北地区，所在地为东北地区的育龄妇女生育更多子女的可能性更低，所在地区为中南地区的育龄妇女生育更多子女的可能性更高；相对于 2019 年，2021 年育龄妇女生育更多子女的可能性更高。

Table 8. Results of ordered logistic regression

表 8. 有序 logistic 回归结果

自变量	OR	Std. Err.	z	P >  z	[95% Conf. Interval]	
受教育程度(参照组小学及以下)						
初中	0.551	0.050	-6.600	0.000	0.462 0.658	
高中/中专/职高	0.390	0.046	-8.060	0.000	0.310 0.490	
本科/大专	0.256	0.035	-9.890	0.000	0.196 0.336	
研究生	0.332	0.129	-2.840	0.004	0.155 0.710	
家庭人均收入	0.872	0.030	-3.960	0.000	0.815 0.933	
家庭规模	1.929	0.059	21.530	0.000	1.817 2.048	
家庭相对收入	1.120	0.040	3.200	0.001	1.045 1.201	
养老保险	0.757	0.061	-3.460	0.001	0.646 0.886	
医疗保险	0.984	0.086	-0.180	0.857	0.829 1.169	
年龄	0.939	0.012	-5.050	0.000	0.916 0.962	
性别(对照组: 男性)	1.015	0.077	0.200	0.843	0.875 1.178	
婚姻状况(对照组: 初婚有配偶)						
再婚有配偶	0.993	0.146	-0.040	0.964	0.745 1.326	
离婚	0.463	0.092	-3.870	0.000	0.314 0.684	
丧偶	2.206	0.748	2.340	0.020	1.136 4.286	

## Continued

个人收入	0.941	0.009	-6.600	0.000	0.925	0.958
城乡(对照组: 城市)	1.453	0.116	4.680	0.000	1.243	1.699
所在地区(对照组: 华北)						
东北	0.300	0.054	-6.650	0.000	0.210	0.428
华东	0.971	0.118	-0.240	0.807	0.765	1.232
中南	1.559	0.192	3.600	0.000	1.224	1.985
西南	1.005	0.138	0.040	0.970	0.768	1.315
西北	1.278	0.211	1.480	0.138	0.924	1.767
年份(对照组: 2019 年)	1.289	0.098	3.350	0.001	1.111	1.495
/cut1	-2.207	0.693			-3.565	-0.849
/cut2	1.025	0.693			-0.332	2.383

## 5.1.2. 稳健性检验

将有序 logistic 回归模型中的因变量——生育行为分组(1 个及以下、2 个、3 个及以上)替换成连续变量(0, 1, 2, ..., 12), 使用 Poisson 回归进行稳健性分析, 得到表 9 所示结果, 其与有序 logistic 回归结果保持一致, 主要解释变量中除医疗保险之外均显著, 故认为生育行为影响因素分析的结果是稳健的。

Table 9. Results of Poisson regression

表 9. Poisson 回归结果

变量	OR	Std. Err.	z	P >  z	[95% Conf. Interval]
受教育程度(参照组小学及以下)					
初中	0.897	0.019	-5.090	0.000	0.860 0.935
高中/中专/职高	0.824	0.022	-7.110	0.000	0.782 0.869
本科/大专	0.762	0.024	-8.760	0.000	0.717 0.810
研究生	0.787	0.064	-2.920	0.003	0.671 0.924
家庭人均收入	0.986	0.003	-4.180	0.000	0.979 0.992
家庭规模	1.097	0.015	6.650	0.000	1.068 1.128
家庭相对收入	1.044	0.009	4.810	0.000	1.026 1.062
养老保险	0.934	0.016	-3.940	0.000	0.903 0.966
医疗保险	1.005	0.019	0.250	0.803	0.968 1.043
年龄	0.989	0.003	-4.020	0.000	0.984 0.994
性别(对照组: 男性)	0.982	0.017	-1.070	0.284	0.949 1.016
再婚有配偶	0.980	0.033	-0.600	0.550	0.918 1.046
离婚	0.786	0.039	-4.910	0.000	0.714 0.865
丧偶	1.087	0.082	1.110	0.267	0.938 1.260
个人收入	0.996	0.002	-2.040	0.042	0.991 1.000
城乡(对照组: 城市)	1.087	0.019	4.920	0.000	1.052 1.124
所在地区(对照组: 华北)					
东北	0.783	0.028	-6.900	0.000	0.730 0.839
华东	0.980	0.023	-0.860	0.391	0.937 1.026
中南	1.081	0.026	3.260	0.001	1.032 1.133
西南	1.013	0.027	0.500	0.616	0.962 1.068
西北	1.041	0.032	1.310	0.189	0.980 1.107
年份(对照组: 2019 年)	1.050	0.013	3.830	0.000	1.024 1.076

## 5.2. 育龄人群生育行为影响因素特征重要性排序

将调查样本按照 4:1 的比例划分训练集和测试集,对调查样本的生育行为进行分类和预测。使用 F1 值和 AUC 作为评价指标, F1 值为模型准确率和召回率的调和平均值, AUC 同时考虑了分类器对于正例和负例的分类能力,在样本不平衡的情况下,依然能够对分类器作出合理的评价。两个评价指标都可以综合评价机器学习算法的性能优劣和稳定性。使用逻辑回归(LR)、随机森林(RF)、XGBoost (XGB)、LightGBM (LGB)四种分类算法对样本进行分类,各分类器的 F1 值和 AUC 如表 10 所示,比较得到,无论是 F1 值还是 AUC 值,LightGBM 算法表现得均较高。因此本文选用 LightGBM 分类算法对样本进行分类。

Table 10. Evaluation indexes of each classifier algorithm

表 10. 各分类器算法评价指标

评价指标	LR	RF	XGB	LGB
F1	0.663	0.702	0.693	0.712
AUC	0.747	0.777	0.769	0.784

由于变量量纲或赋值的差异,回归模型的系数大小并不能直接代表单个变量影响生育行为的重要程度。故使用 LightGBM 分类算法对各解释变量进行特征重要性排序,如图 3 所示,影响育龄人群生育行为的特征按照重要性由大到小排序分别为家庭规模、个人收入、家庭相对收入、家庭人均收入、年龄、所在地区、受教育程度、婚姻状况、养老保险、城乡、性别、医疗保险和调查年份。可见,相较于个人的受教育程度以及是否购买养老保险等特征,家庭所能提供的生活照料以及个人、家庭的经济状况是影响育龄人群生育行为更为重要的因素。这也提示我们要以帮助改善家庭生育条件为发力点,通过提供配套生育支持,保证育龄妇女的就业权益等来鼓励育龄人群产生更多的生育行为。

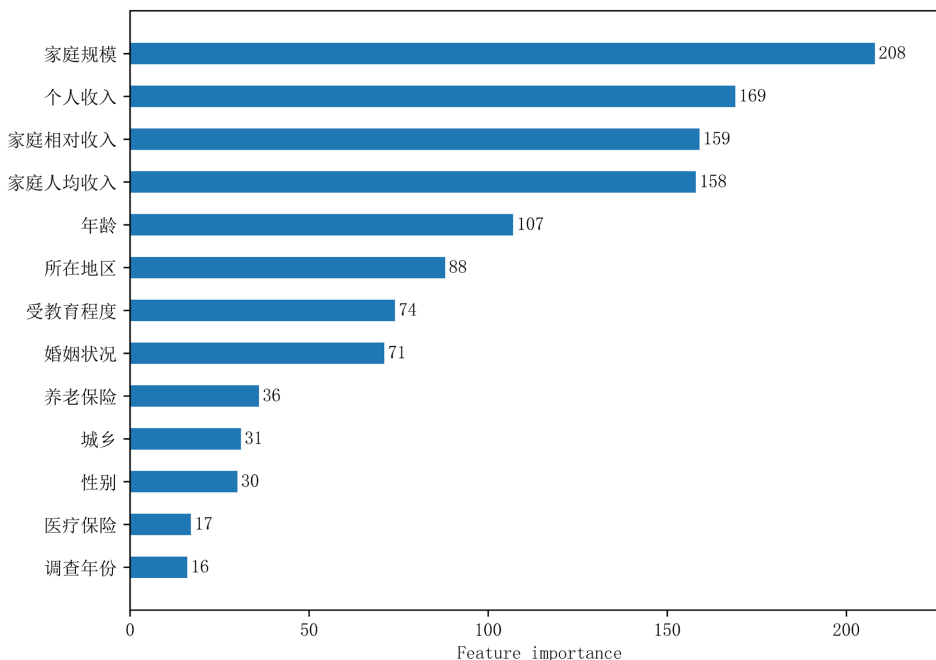


Figure 3. Importance of characteristics of influencing factors of reproductive behavior

图 3. 生育行为影响因素特征重要性

### 5.3. 生育意愿对生育行为中介效应分析

意愿和行为通常都是一个内部的体系，特别是当个体受到外界因素的影响之后，它们通常会通过意愿的转变来做出反应，所以，在对生育行为进行研究时，必须将生育意愿的影响全面地考虑进去[10]。使用中介效应模型探究上述显著的社会经济相关变量，是否存在中介机制，即通过影响生育意愿从而影响生育行为，如果存在中介效应，则具体的中介效应占比为多大，中介效应检验的路径图4如下：

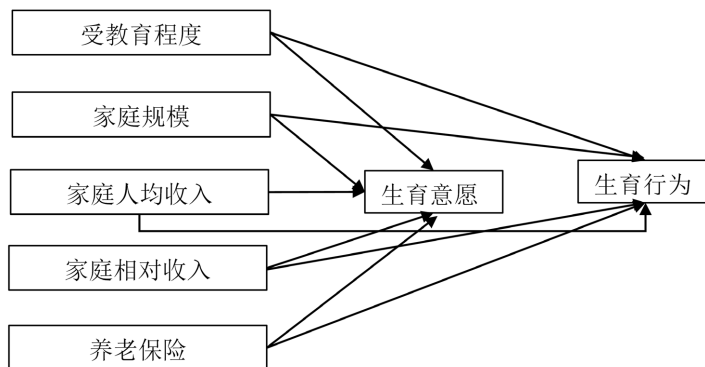


Figure 4. Test path diagram of mediation effect of reproductive willingness on reproductive behavior  
图 4. 生育意愿对生育行为的中介效应检验路径图

中介效应的模型结果如表 11 所示，受教育程度和家庭规模的中介效应显著，家庭人均收入、家庭相对收入、养老保险的中介效应不显著。受教育程度对生育行为影响的直接效应为-0.106，中介效应为-0.006，中介效应占比 5.36%；家庭规模对生育行为影响的直接效应为 0.156，中介效应为 0.015，中介效应占比为 8.77%。

Table 11. Results of the mediation effect model  
表 11. 中介效应模型结果

	生育行为			生育意愿			生育行为		
	系数	t	P	系数	t	P	系数	t	P
受教育程度	-0.106	-9.737	0.000	-0.032	-2.478	0.013	-0.100	-9.415	0.000
家庭规模	0.156	24.393	0.000	0.084	10.922	0.000	0.141	22.208	0.000
家庭人均收入	0.000	-5.171	0.000	-0.122	-0.077	0.939	0.000	-5.276	0.000
家庭相对收入	0.042	3.621	0.000	0.023	1.627	0.104	0.038	3.348	0.001
养老保险	-0.070	-3.317	0.001	-0.020	-0.804	0.422	-0.067	-3.219	0.001
生育意愿							0.179	13.184	0.000
控制变量		是			是			是	
R <sup>2</sup>		0.308			0.068			0.340	
调整 R <sup>2</sup>		0.306			0.064			0.337	
F	F(13, 3605) = 123.529, P = 0.000			F(13, 3605) = 20.26, P = 0.000			F(14, 3604) = 132.62, P = 0.000		

### 5.4. 不同模型结果比较分析

有序 logistic 回归结果显示，主要解释变量中，受教育程度、家庭人均收入、家庭规模、家庭相对收入、养老保险对生育行为影响显著，而医疗保险对生育行为影响不显著；控制变量中，除了性别，剩余

的年龄、婚姻状况、个人收入、城乡、所在地区、年份均对生育行为影响显著。接着,运用 Poisson 回归进行稳健性检验,检验结果与有序 logistic 回归结果一致,故认为生育行为影响因素模型是稳健的。有序 logistic 回归和 Poisson 回归结果均显示,家庭规模和城乡是影响生育行为的重要因素,而 LightGBM 模型中,特征重要性排在前二的是家庭规模和个人收入。中介效果也显示,家庭规模和受教育程度会对生育行为产生中介效应。可见,在不同模型中,家庭规模是生育行为的共同影响因素。

## 6. 结论与建议

### 6.1. 结论

生育水平对于优化人口结构、促进人口均衡发展具有重要意义。本文基于中国社会状况综合调查(CSS)数据,运用有序 Logistic 回归、Poisson 回归、LightGBM 模型、中介效应模型等方法,对不同特征育龄群体生育行为进行差异分析并探究影响生育行为的社会经济因素,得到结论如下:

1) 从生育行为来看,当前育龄群体选择终身生育 1 个或不生孩子的样本量占比 41.11%,选择终身生育 2 个孩子及以上的样本量占比为 58.89%。除性别外,不同特征育龄群体之间生育行为差异显著。

2) 有序 Logistic 回归结果显示,生育行为受家庭人均收入、受教育程度、家庭相对收入、家庭规模、养老保险的显著影响,家庭人均收入越高、受教育水平越高、家庭相对收入越低、家庭规模越小、参与养老保险的育龄群体其生育更多子女的可能性更低。

3) 运用 LightGBM 模型对影响调查样本生育行为的特征重要性排序,从大到小分别为,家庭规模、个人收入、家庭相对收入、家庭人均收入、年龄、所在地区、受教育程度、婚姻状况、养老保险、城乡、性别、医疗保险和调查年份,相较于个人的受教育程度以及是否购买养老保险等特征,家庭所能提供的生活照料以及个人、家庭的经济状况是影响育龄人群生育行为更为重要的因素。

4) 中介效应模型结果显示,受教育程度、家庭规模通过影响生育意愿,从而对生育行为产生影响,两者的中介效应占比分别为 5.36%、8.77%。

### 6.2. 建议

基于上述研究结论,本文提出以下几点建议:

1) 完善社会公共服务供给,提高相对收入,缓解家庭生育养育教育负担。完善教育、医疗、环境、公共安全各方面的社会公共服务供给,一定程度上降低育龄人群的绝对支出,提高其相对收入,从而提高育龄人群增加生育的概率。要大力发展普惠制的公共托育场所等福利设施,加快制定行业标准和管理体系,推动优化托幼服务业的发展,降低年轻一代的生育成本,让那些缺少充足家庭支持的育龄群体能够免去孩子生下来无人照看的后顾之忧。同时,要加强养老服务专业设施建设,以“一老一小”为重点完善人口服务体系。

2) 构建生育友好型社会,引导监督用人单位保障婚育女职工各项合法权益。受教育程度高的育龄群体生育更多子女的可能性较小,尤其是对于女性来说,受教育水平高会通过提高机会成本从而降低生育水平,因此政府应根据实际情况建立育儿措施,推进弹性工作制,实施育龄人群的带薪(陪)产假等制度来缓解育龄群体在家庭和工作之间的矛盾,保障育龄妇女的就业合法权益,对以生育为由解聘的单位要严格惩治,鼓励弹性制、远程办公等灵活的工作模式,对聘用多孩妇女用人单位予以减税和补贴等。

3) 加强婚育新风引导,形成婚育友好氛围。要倡导新时代婚育文化,通过“婚育新风进万家”活动,将科学文明的婚育观传播到、影响到千家万户,形成生育友好型的社会氛围。要倡导适龄生育、适当缩小生育间隔、优生优育等,要鼓励夫妻共同承担家庭育儿责任,落实“父母同育”的鼓励政策,增加走访和宣传工作,开展“能干奶爸”、“优秀生育家庭”等评选活动,促进儿童健康成长、家庭幸福和谐。

## 参考文献

- [1] 国家统计局. 中华人民共和国 2022 年国民经济和社会发展统计公报[EB/OL]. [http://www.stats.gov.cn/sj/zxfb/202302/t20230228\\_1919011.html](http://www.stats.gov.cn/sj/zxfb/202302/t20230228_1919011.html), 2023-02-28.
- [2] 国家统计局. 2020 年全国人口普查资料[EB/OL]. <http://www.stats.gov.cn/sj/pcsj/rkpc/7rp/indexch.htm>, 2021-07.
- [3] 晏月平, 张舒贤. 多层社会资本对育龄群体生育行为的影响——基于 CFPS 数据库的实证研究[J]. 河南社会科学, 2023, 31(3): 113-124.
- [4] 宋月萍, 宋正亮. 生育行为对老年女性健康的影响[J]. 人口研究, 2016, 40(4): 76-87.
- [5] 刘璐婵, 张瑞利. “全面三孩”政策背景下教育成本的生育抑制效应[J]. 统计与决策, 2022, 38(13): 83-87.
- [6] 冯永琦, 于欣晔. 家庭投资风险偏好对生育行为的影响研究——基于 CGSS2017 数据的分析[J]. 经济问题, 2022(7): 54-62.
- [7] 李涛, 方明, 伏霖, 等. 客观相对收入与主观经济地位: 基于集体主义视角的经验证据[J]. 经济研究, 2019, 54(12): 118-133.
- [8] 王一帆, 罗淳. 促进还是抑制? 受教育水平对生育意愿的影响及内在机制分析[J]. 人口与发展, 2021, 27(5): 72-82+23.
- [9] 刘倩, 肖梅, 黄洪滔, 等. 基于 LightGBM 模型的公交线路串车状态识别方法[J]. 交通信息与安全, 2022, 40(5): 102-111.
- [10] 梁宏. 从生育意愿到生育行为: “全面两孩”政策背景下二孩生育决策的影响因素分析[J]. 南方人口, 2018, 33(2): 1-14.