

我国人口出生率影响因素研究

王 英

贵州大学数学与统计学院, 贵州 贵阳

收稿日期: 2022年3月25日; 录用日期: 2022年4月30日; 发布日期: 2022年5月7日

摘 要

“单独二孩”提出以来, 人口出生率再一次成为广泛关注的焦点。数据显示, 自1987年以来我国出生率持续降低, 人口出生率的下降必然使得我国人口老龄化问题加剧, 也会给社会以及国家的发展都带来一定的不良影响。经查阅得知: 经济、医疗条件、教育及国家颁布的生育政策等方面均对我国的基本人口国情存在影响。本文旨在通过R软件建立回归方程及对所得回归数学模型进行分析, 以改革开放以来中国人口基本国情为基础, 对全国人口出生率影响因素进行探究。经试验分析得知: 老年抚养比、居民消费水平、商品房平均销售价格及参加生育保险人数均对出生率具有显著的影响, 其中, 最显著的是老年抚养比和居民消费水平。

关键词

人口基本国情, 主成分回归分析法, 逐步回归分析法, 出生率

Study on the Influencing Factors of Birth Rate in China

Ying Wang

School of Mathematics and Statistics, Guizhou University, Guiyang Guizhou

Received: Mar. 25th, 2022; accepted: Apr. 30th, 2022; published: May 7th, 2022

Abstract

Since the “single second child” was proposed, the population birth rate has once again become the focus of widespread attention. Data show that since 1987, China’s birth rate has continued to decline, the decline in the population birth rate will inevitably make China’s population aging problem aggravated, but also to the development of society and the country will bring certain adverse effects. After consultation, it is learned that the economy, medical conditions, education and the birth policy promulgated by the state have an impact on the basic population and national condi-

tions of our country. This paper aims to establish the regression equation through R software and analyze the mathematical model of the obtained regression, and to explore the influencing factors of the national population birth rate based on the basic national conditions of the Chinese since the reform and opening up. After experimental analysis, it is learned that the old-age dependency ratio, the consumption level of residents, the average sales price of commercial housing and the number of people participating in birth insurance all have a significant impact on the birth rate, of which the most significant is the elderly dependency ratio and the consumption level of residents.

Keywords

Basic Population Conditions, Principal Component Regression Analysis Method, Stepwise Regression Analysis Method, Birth Rate

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

1.1. 研究背景及研究意义

2016年1月1日我国实施全面放开二孩的政策以来,人口出生率并没有迎来预期的爆发式增长。《中国统计年鉴(2017~2021)》公布的数据表明,2016~2021年全年人口出生率逐年下降,其中2016年人口出生率为12.95%,2017年为12.43%,2018年为10.94%,2019年为10.48%,2020年为8.52%。2020年11月3日发布的《中共中央关于制定国民经济和社会发展第十四个五年规划和二〇三五年远景目标的建议》指出“优化生育政策,增强生育政策包容性”,这表明十四五期间全面放开生育政策成为一种可能。人口出生率是一个国家立国之根本,是决定未来国家全面发展速度的重要指标。人口问题不只是中国存在,是全世界几乎每一个国家都不得不待解决的国家发展的重要问题,也是我国亟待解决的问题,人口问题与国家的经济发展密切相关。在影响出生率的众多因素中,国家政策是一个重要影响因素,例如奖励政策,日本与泰国政府在此方面就做的比较好。也可以通过宗教信仰对人口出生率产生一定的影响,每一个国家都需有宗教的存在,如果宗教信仰反对避孕,这个国家的出生率可以得到保证。对出生率的影响因素进行研究分析有利于政府制定生育政策,节约落实国家生育政策的人力以及物力,促进人口结构优化等。人均资源不足,生存空间狭小,进而使得我国改革开放以来居民生活水平提升缓慢,即使进入21世纪,生活水平有显著提升,但是相对于一些发展较快的国家而言,我们国家因人口问题未能妥善解决而导致落后于其他发达国家。人口老龄化严重、社会保障制度残缺、不健全这些都对我国的经济可持续发展产生较大的负面影响。城乡人口就业问题不能良好解决,劳动力市场就业机制不完善,导致现有劳动力得不到充分利用。同时,我国人口老龄化形式正在加剧,这就意味着国家必须将更多的资金用于对老年人社会生活的保障,从而投入到国家经济建设发展的资金就会相应减少,不利于国家经济的发展[1]。其次,人口老龄化意味着青壮年的比例减小,青壮年群体是一个国家发展进步最主要的劳动力,劳动力的比例减小,必然导致整个市场劳动力紧缺,从而市场产量低下,将直接影响到国家经济的持续发展。

增加人口基本数量,提升大众居民素质水平,在我国可持续发展战略实施中起着关键性作用。众所周知,人口出生率就是人口增长的宏观表现,随着出生率的增长人口也必然会增长。长期来看,人口出

生率是由政策和经济发展综合影响的。在发展前期,中国人口增速过快,此时中国主要通过实施计划生育政策来降低人口增速。计划生育政策在生育率下降过程中所起的作用可以分为两类:一类是政策对生育行为的直接约束;另一类则是有关的宣传教育通过对人们生育意愿的影响进而改变了生育行为。“上世纪七八十年代中国生育率的异常下降,不仅有强制性因素,也有重要的自愿性因素”(陈卫,2005) [2]。但在当前情况下,随着“单独二孩”到“全面二孩”的政策实施,相关调查显示,目前独生子女家庭生育意愿很低,半数以上的家庭不愿意生育二胎,并且发达地区生育意愿普遍低于一般地区,可见政策对生育行为的直接约束已显著弱化,即使政策鼓励多生,我们国家近几年来人口出生率仍呈现出逐年下降的趋势。社会经济发展对生育意愿和人口出生率的影响正日益突出,随着经济的快速发展,养育成本过高,育龄人群的生活以及经济负担加重,这是很多家庭“不敢生、不愿生”的主要原因[3]。因此本文选择从国家的社会经济发展出发探索影响出生率的主要影响因素就有了理论支撑与实际意义。

1.2. 文献综述

自中国出生率备受关注以来,中国实施了多项政策用以提高人口出生率。在“全面两孩”实施以后,庄亚儿等(2021) [3]在研究中得出结论 80 后生育人群平均打算生育为 1.82,且她们中相当一部分人仍处于生育旺盛期,所以实施“全面两孩”政策后,其生育意愿得到释放,从而该部分妇女打算生育子女数在不同出生年代中是最高的。90 后平均打算生育子女数仅为 1.70,并且目前还未生育妇女的平均打算生育子女更低,两者均低于 1.76 的总体水平。受社会经济因素、婚姻状况以及不孕不育等因素的影响,实际生育子女数一般都低于打算生育子女数。所以在社会经济支持政策没有达到足够强度的情况下,未来我国的出生率仍会下降。在探索出生率或者生育率影响因素的文献中,刘金菊(2020) [4]研究中国城镇妇女的生育成本时,发现中国城镇妇女每生育一个孩子都会损失数十万不同的经济收入,对于受过高等教育的女性,其损失更为严重。生活在北京、上海等发达地区的女性,其工作时间的损失会略高于全国城镇女性,并且其损失的收入也会高于全国城镇女性,很好的解释了为什么发达地区的出生率会更低以及随着国家 GDP 的增长出生率也会下降的问题。生育推迟是导致出生率下降的一个重要原因,胡佩等(2020) [5]在研究房价与生育延迟的关系时,得出的结论为房价上涨将显著延迟初生时间,即房价的涨幅与延迟年限具有正向的关系。当房价上涨 1%时,就会延迟约 1.05 年的时间,并且家庭收入不同延迟的年限也会有所不同,其中对中档收入家庭的影响最为显著。在杨龙见,陈建伟等(2013) [6]一文中使用 GMM 估计方法研究政府教育支出对出生率是否有影响,结果发现,在一定的条件下,政府教育支出会降低人口出生率。在探究影响出生率的主要影响因素时,李松臣,张世英(2008) [7]两人以 31 个省份的数据为研究对象,选取了收入水平等多个变量作为解释变量,出生率为被解释变量,运用逐步回归的方法研究得出人口组成以及收入水平对出生率有影响。杨慧诗,孟晗(2010) [8]在探究世界各国人口出生率的影响因素时,以世界 27 个国家为研究对象,使用逐步回归分析法对其分析,得出结论青少年抚养比和婴儿死亡率是主要影响因素,并得出回归方程。李建伟(2014) [9]研究了经济因素、社会保障程度、养育成本等与出生率或者生育率的关系,该三种因素均与出生率表现出负向的关系,其中经济因素具有重要影响,并指出了我国育龄妇女生育率的发展趋势。陈阔(2015) [10]使用双对数线性回归模型得出结论养老保险与人口出生率有正相关的关系。刘丽萍(2018) [11]采用 lasso 回归法研究 GDP、人均薪酬等对出生率的影响,得出结论,国家 GDP、人均薪酬、少儿抚养比均对出生率产生正向影响。玛依拉·吐尔逊,常轩(2019) [12]二人基于动态 GMM 模型和门限模型对出生率影响因素进行分析,发现经济发展对出生率产生的抑制作用是由生育保险实现的。田青(2020) [13]以 2007 年至 2017 年十年的相关数据为研究对象,其中包含了二十个变量,基于 OMP 算法对出生率影响因素进行探究,最后得出影响人口出生率最为显著的变量为少年抚养比、人均薪酬以及居民消费价格指数。在对人口出生率的预测文献中,孙鑫鑫(2020) [14]等人 1949

年~2017年的国家出生率为研究对象,对其建立时间序列模型,从而预测2018年~2022年五年的中国人口出生率,得出的五年的出生率均高于12%,但其预测的2018年以及2019年的数据与实际数据相差均高于1%,可见其预测结果并不可靠。张炜和朱家明(2017) [15]对人口组合进行预测时运用的方法是ARIMA模型以及二次指数平滑法。并且由张现苓等研究者(2020) [16]预测可知,21世纪上半叶中国人口问题将持续加剧,2027年后将进入负增长模式,并且适龄劳动人口进入负增长的时间会早于总人口,负增长的速度也会快于总人口。

由以上文献可得,在计划生育的实施至未来一段时间内,国家生育政策会对出生率产生显著的影响,但是随着国家经济的不断发展,国家生育政策的影响力逐渐减弱,社会经济以及其它因素的影响越来越大,并且在未来的一段时间内中国的人口问题将会加剧。本文在已有研究文献的基础上,选取已有文献所探索出来的对人口出生率具有显著影响的变量作为解释变量,将出生率作为被解释变量,首先对所得数据进行描述性统计,根据数据特有的属性选择合适的回归方法对其建立全模型回归分析,并使用逐步回归分析在这几个显著影响的变量中寻找主要影响变量,给出显著变量的回归方程。

1.3. 研究方法

1.3.1. 文献研究法

该方法的优点是可以超越时间以及空间的限制,通过对不同时期的文献进行调查去研究及其广泛的社会情况。不仅如此,文献研究法还可以避免实地调查研究的种种误差,可以节约大量的人力以及物力,是一种高效率的调查方法。本文根据研究内容,参考有关出生率影响因素探究的文献,从中探索出对出生率产生显著影响的因素,将其作为本文探究的解释变量,为本文的变量选取提供了充分的理论以及现实依据。并通过参考已有文献的理论依据及方法经验从而为本文的因素分析和建模提供了探索方向及分析方法。

1.3.2. 主成分回归分析法

该方法的思想是降维,可以在保留原始数据大部分信息的情况下将原始变量转换为几个不具有共线性的综合变量进行建模分析。解释变量之间存在严重的共线性,若直接使用最小二乘法进行建模分析,得出的结果不具有说服力。因此本文转用主成分回归分析的方法,对所选取的影响因素选取主成分,最后通过最小二乘的方法对选取的主成分进行回归建模,从而可以达到消除变量之间的多重共线性再建模的目的,得出的模型更加可靠,更具有理论依据。

1.3.3. 逐步回归分析法

逐步回归分析法可以剔除不重要并且引起严重多重共线性的解释变量,最后输出AIC最低的最优模型。主成分回归输出结果显示,输出方程的R平方仅有0.1335,从而可知该方程的拟合度较差,建立的全模型并不理想。因此本文又转用逐步回归分析的方法探究在所选取的解释变量组合中消除多重共线性后对人口出生率具有显著影响的变量,得到降低多重共线性后对出生率产生主要影响的变量,最后得到出生率与显著影响变量之间的回归方程。

2. 数据来源、现状描述和变量选取

2.1. 数据来源

本次研究数据均来自《中国统计年鉴》,统计年鉴作为一部覆盖国民经济和社会发展各方面的大型权威工具书,其中的数据真实可靠,增强了本次研究的可靠性以及真实性,让所得结论更具有说服力。根据本文的研究内容,对统计年鉴中的原始变量进行了收集整理,得出2001年至2020年的中国人口出

生率数据以及每个影响变量的数据进行研究分析，共选取了九个影响因素作为解释变量，将出生率作为被解释变量。

2.2. 现状描述

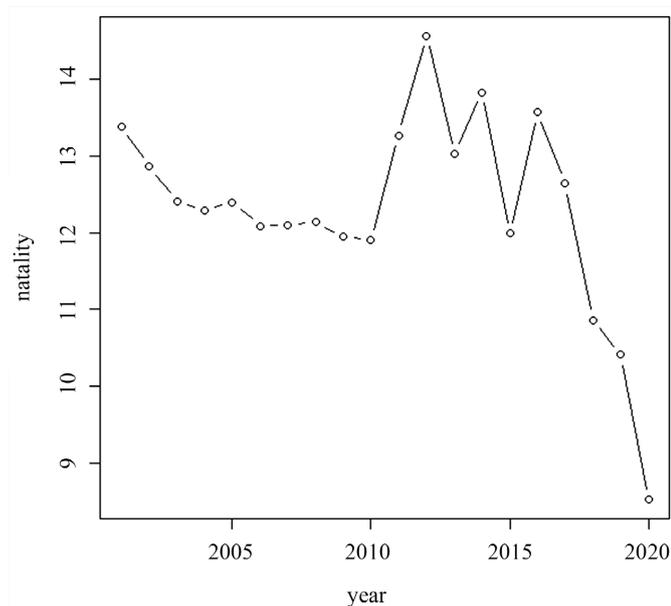


Figure 1. Trends in the birth rate from 2001 to 2020

图 1. 2001 年至 2020 年出生率走势

由图 1 可知，从 2001 年到 2010 年中国人口出生率一直处于一种下降的趋势。众所周知，中国的人口基数庞大，为了防止中国人口的剧增，1971 年实施了计划生育，在实施计划生育期间，部分生育人群自愿少生以及国家严抓严防，采取一系列的行动有效的减少了出生率，所以 2001 年到 2010 年的出生率才会一直处于下降的状态，同时中国人口也慢慢进入老龄化，为了缓解老龄化的问题，在 2011 年初，中国颁布了计划生育新政策，满足具体情况的家庭可以生育二胎，并于同年 11 月，各地全面实施双独二胎政策。2013 年末，为了再一次提高出生率，中国实施单独二孩政策，从而使得 2011 年至 2014 年出生率有所上升。由于 2015 年是羊年，部分人受传统文化的影响不太喜欢养宝宝，并且在 2015 年部分地区还没有开始实施二胎政策等原因，所以 2015 年是 2011 至 2016 年之间出生率下降的唯一一年。2016 年 1 月 1 日，中国各地全面实施二孩政策，从而导致 2016 年的出生率有所上升，又因为 2015 年的出生率有所下降，相比之下 2016 年便成为了增速最快的一年，尽管如此，其上升的幅度并没有达到期望值。从 2017 年开始出生率又处于下降的趋势，并且其下降的幅度远大于在 2011 年至 2016 年的上升幅度，在这四年中直接从 12.95 下降到 8.52。通过两孩政策的全面实施，可以看到较为显著的积极效应，人口呈现出一个良好态势，但由于受育龄妇女规模下降、婚育年龄推迟等无法短时间改善的因素影响，经济社会因素对人们生育意愿和生育行为的影响更为明显。由本文文献综述部分可知：养育成本、托育服务等众多因素，导致育龄人群不愿意多生育孩子。

2.3. 变量选取

本文变量的选取是基于已有文献进行的，共选取了九个变量，这九个变量涉及到已有文献中对出生率具有显著影响的多个方面例如人口组成、国家的经济发展水平、国家总体的受教育水平以及生育保险

等，具体如下。

2.3.1. 人口因素

自从中国实施计划生育以来，中国的人口结构在不断的发生变化。计划生育实施初期，中国人口数量庞大，此时中国的适龄劳动人群占比高。但现如今，由于之前计划生育政策等影响，中国的适龄劳动人口占比已经逐渐降低，但老年抚养却在增高，从而也就增大了现代适龄劳动人口的经济压力，这也就成为了现代适龄生育人群不愿生养的重要原因。刘岩(2020) [17]在研究人口年龄结构与经济发展的关系时发现，老年抚养比和少儿抚养比直接影响了我国未来经济增长的变化。而经济发展是对中国出生率的重要影响因素，所以本文根据已有文献的结论选取了少儿抚养比以及老年抚养比作为此次探究的因素。少儿抚养比的计算方式为： $CDR = (0\sim 14 \text{ 岁少年儿童人口数} / 15\sim 64 \text{ 岁劳动年龄人口数}) * 100\%$ ，其中：CDR为少年儿童抚养比。由计算公式可初步判断，少儿抚养比越高，表示每一百名劳动年龄人口要负担的少年儿童就越多，劳动人口负担可能就越重，从而直接影响到了人口出生率的高低。老年抚养比的计算方式为： $ODR = (65 \text{ 岁以上人口数} / \text{劳动年龄人口数}) * 100\%$ ，其中：ODR为老年抚养比。同样由计算公式可初步判断，当老年抚养比越高时，适龄劳动年龄人口的负担越重，从而可能会导致出生率下降。

2.3.2. 经济因素

通常来说经济发展得越好，人民的生活水平就会越高，劳动人口就能赚取更多的资金去抚养少年儿童，从而出生率就会增加，反之则降低。但由世界各国的数据来看，随着经济的不断发展，各国居民的生育意愿都在降低，各国的出生率均处于下降的状态，特别是高收入的国家，其出生率远远低于低收入的国家。本文选取影响出生率的经济因素有国内生产总值和居民消费水平。

国内生产总值(GDP)反映了国家经济发展状况，国家的经济发展得越好，GDP越高，随着国家GDP的提高，交通越来越便利，越来越多的人走出自己的家乡，去接触新鲜事物，视野变得更开阔，从而导致了一部分适龄人群有“乘年轻，多玩几年”的想法，追求自由无约束的生活。还在一定程度上改变了生育观念，致使一部分人向往单身生活，成为不婚主义者。并且随着GDP升高，适龄人群生育子女的损失也会随之增加，从而GDP对人口出生率就产生一定的负面影响。

居民消费水平可以用来衡量一个国家居民的总体经济发展状况以及人民的生活水平，如果居民消费水平有所提高，便可知人民的生活水平也在进一步的改善，即更多人有经济实力去追求更好的生活质量。居民消费水平反映出人们的生活水平，居民的生活水平的高低也会或高或低的影响着人口出生率，从而居民消费水平与出生率也就产生了必然的联系。

2.3.3. 教育水平

教育水平是指人们接受教育的程度，由刘金菊(2020)一文知，受教育程度越高的女性，生育孩子付出的代价就越高，在经济方面的损失也会高于平均女性生育孩子的损失。并且受教育水平普遍提高之后，人们对生育子女的态度会有所改变，女子也不再安定于在家相夫教子，做一个家庭主妇，而是更多的会选择有一份自己安定的工作，有一份自己稳定的收入，更多的依靠自己去赚钱，做一个独立的女性。从而导致人们的生育观念不再是多子多福，而是晚育，少生，优生。这种观念的转变，对出生率产生了很大的负面影响，从而导致出生率持续下降。所以随着国家教育水平的提高，越来越少的女性愿意多生，即受教育水平与出生率具有负向的关系。本文用来反映受教育程度的变量为本科招生数(万人)。

2.3.4. 其它变量

在相关文献中，我们了解到随着房价的增加会延迟女性的首次生育时间。在这个充满竞争的社会，不论是男生还是女生，为了让自己能在更好的城市站稳脚跟，都会选择先努力工作买一套属于自己的房

子, 然后再考虑成家的问题。然而若女性超过 35 岁首次妊娠就是大龄产妇, 大龄产妇的危害可想而知, 所以若女性因为房价延迟了首次受孕时间, 则其为了自生的身体健康, 或者说为了优生, 其必然会选择少生, 出生率也会因此降低, 所以本文将商品房平均销售价格(元/平方米)作为其中一个解释变量。又由相关文献知生育保险, 婴儿死亡率以及离婚率均对出生率具有显著影响, 所以本文又将参加生育保险人数(万人)、婴儿死亡率(%)以及粗离婚率(%)作为其中的三个解释变量。

3. 基于主成分回归分析对指标进行全模型建模分析

3.1. 数据处理

为了方便处理数据, 本文将所有的变量使用相应的变量符号代替, 具体如表 1, 具体数据如表 2。

Table 1. Variables
表 1. 变量

变量名	变量
X1	少儿抚养比(%)
X2	老年抚养比(%)
X3	国内生产总值(亿元)
X4	居民消费水平(元)
X5	粗离婚率(%)
X6	普通本科招生数(万人)
X7	婴儿死亡率(%)
X8	商品房平均销售价格(元/平方米)
X9	参加生育保险人数(万人)
Y	中国人口出生率(%)

并由数据之间量纲不同的特性将数据进行标准化, 数据标准化可以避免由于量纲的不同导致结果出现偏差的问题, 并且标准化后的数据具有可比性, 是在数据统计分析中经常被研究者运用的一种方法。并进一步对数据进行了描述性统计以及相关分析, 从而初步了解数据的基本特性, 为之后的影响因素研究奠定了基础。

Table 2. Data on relevant variables from 2001 to 2020
表 2. 2001 年至 2020 年相关变量数据

年份	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9
2001	13.38	32.0	10.1	110863.1	3968	0.98	138.1835	30.0	2170	3455.1
2002	12.86	31.9	10.4	121717.4	4270	0.90	158.7939	29.2	2250	3488.2
2003	12.41	31.4	10.7	137422.0	4555	1.05	182.5262	25.5	2359	3655.4
2004	12.29	30.3	10.7	161840.2	5071	1.28	209.9151	21.5	2778	4383.8
2005	12.4	28.1	10.7	187318.9	5688	1.37	236.3647	19.0	3168	5408.5
2006	12.09	27.3	11.0	219438.5	6319	1.46	253.0854	17.2	3367	6458.9
2007	12.10	26.8	11.1	270092.3	7454	1.59	282.0971	15.3	3864	7775.3
2008	12.14	26.0	11.3	319244.6	8505	1.71	297.0601	14.9	3800	9254.1

Continued

2009	11.95	25.3	11.6	348517.7	9249	1.85	326.1081	13.8	4681	10875.7
2010	11.90	22.3	11.9	412119.3	10575	2.00	351.2563	13.1	5032	12335.9
2011	13.27	22.1	12.3	487940.2	12668	2.13	356.6411	12.1	5357	13892.0
2012	14.57	22.2	12.7	538580.0	14074	2.29	374.0574	10.3	5791	15428.7
2013	13.03	22.2	13.1	592963.2	15586	2.57	381.4331	9.5	6237	16392.0
2014	13.83	22.5	13.7	643563.1	17220	2.67	383.4152	8.9	6324	17038.7
2015	11.99	22.6	14.3	688858.2	18857	2.79	389.4184	8.1	6793	17771.0
2016	13.57	22.9	15.0	746395.1	20801	3.02	405.4007	7.5	7476	18451.
2017	12.64	23.4	15.9	832035.9	22969	3.15	410.7534	6.77	7892	19300.2
2018	10.86	23.7	16.8	919281.1	25245	3.20	422.1590	6.1	8726	20434.1
2019	10.41	23.8	17.8	986515.2	27504	3.36	431.2880	5.6	9310	21417.3
2020	8.52	26.2	19.7	1013567.0	27439	3.09	443.1154	5.4	9860	23567.3

注：有部分数据被进行了四舍五入处理。

3.2. 主成分回归分析的基本理论

参考任雪松, 于秀林(2010) [18]《多元统计分析》可知主成分分析的基本理论如下。

设在某一研究中涉及的解释变量有 p 个, 可用 X_1, X_2, \dots, X_p 表示。于是 p 个指标便可构成一个 p 维的随机向量 $\mathbf{X} = (X_1, X_2, \dots, X_p)'$, 设 $\boldsymbol{\mu}$ 为该随机向量 \mathbf{X} 的均值向量, 方差为 Σ 。

将 \mathbf{X} 进行线性变换后形成新的不相关的综合变量, 用 \mathbf{Y} 表示, 如下式:

$$Y_1 = \mu_{11}X_1 + \mu_{12}X_2 + \dots + \mu_{1p}X_p$$

$$Y_2 = \mu_{21}X_1 + \mu_{22}X_2 + \dots + \mu_{2p}X_p$$

$$\vdots$$

$$Y_p = \mu_{p1}X_1 + \mu_{p2}X_2 + \dots + \mu_{pp}X_p$$

为了更好的效果, 我们通常希望 Y_i 的方差尽可能大并且各 Y_i 之间互相独立。又 $\text{var}(Y_i) = \text{var}(\boldsymbol{\mu}_i' \mathbf{X}) = \boldsymbol{\mu}_i' \Sigma \boldsymbol{\mu}_i$, 而对任意的常数 c , 有 $\text{var}(c\boldsymbol{\mu}_i' \mathbf{X}) = c^2 \boldsymbol{\mu}_i' \Sigma \boldsymbol{\mu}_i$, 此时若对 $\boldsymbol{\mu}_i$ 不加以任何的限制, $\text{var}(Y_i)$ 就可以任意增大, 则研究问题就会没有任何意义, 所以需对线性变换有如下三条约束:

$$1) \boldsymbol{\mu}_i' \boldsymbol{\mu}_i = 1 (i = 1, 2, \dots, p)$$

$$2) Y_i \text{ 与 } Y_j \text{ 互不相关 } (i \neq j, i, j = 1, 2, \dots, p);$$

3) Y_1 是 X_1, X_2, \dots, X_p 的一切满足原则(1)的线性组合中方差最大者, Y_2 是与 Y_1 不相关的 X_1, X_2, \dots, X_p 一切线性组合中方差最大者…… Y_p 是与 Y_1, Y_2, \dots, Y_{p-1} 都不相关的 X_1, X_2, \dots, X_p 的所有线性组合中方差最大者。

在该约束原则下确定的综合变量 Y_1, Y_2, \dots, Y_p 被称为原始变量的第一, 第二……第 p 个主成分。其中, 方差所占比重最大的为第一主成分, 第二主成分次之, 以此类推。在实际工作研究中, 通常只需要选取方差最大的前几个或者一个主成分来简化系统结构, 把握问题的本质。

3.3. 对中国人口出生率进行建模分析

本文根据所选取的变量, 在 R 语言中进行建模分析。

为了初步了解数据的特性，本文首先对还未进行标准化的数据进行初步统计描述分析，如表 3。由表 3 可知，中国人口出生率在近二十年来的最大值为 14.57，最小值为 8.52，方差为 1.310253，波动并不大，从而可知在近二十年来中国人口出生率的变化是一个缓慢下降的过程。但与之形成对比的是，老年抚养比以及少儿抚养比的波动，由表可知，少儿抚养比的方差达到了 3.489609，老年抚养比的方差达到了 2.735998，所以在近二十年来，老年抚养比以及少年儿童抚养比均发生了很大的波动，结合表 2 可知，少年儿童抚养比近年来一直处于下降的趋势，而老年抚养比一直有上升的趋势。并有相关研究表明在 2028 年中国老年抚养比将超过少年抚养比，这也就意味着在其他条件一定的情况下，人口生产线下降，边际人口红利达到负值。在所选取的剩余变量中，方差最大的就是国内生产总值，随着国家的不断发展，国内生产总值得到了很大的改善，从 2001 年的 110863.1 到 2020 年的 1013567，可见国家经济发展的迅速，同时带动着居民消费水平、房价以及参保人数等变量不断的提高。

Table 3. Descriptive statistics for variables

表 3. 变量的描述性统计

变量	中位数	均值	方差	最小值	最大值
Y	12.35	12.31	1.310253	8.52	14.57
X1	24.55	25.65	3.489609	22.1	32.00
X2	12.10	13.04	2.735998	10.10	19.70
X3	450030	486914	302911.8	110863	1013567
X4	11622	13401	8100.908	3698	27504
X5	2.065	2.123	0.8166498	0.9	3.360
X6	353.9	321.7	95.92892	138.2	443.1
X7	12.60	13.99	7.656955	5.4	30.00
X8	5194	5362	2422.989	2170	9860
X9	13114	12539	6688.782	3455	23567

由于本文所选取的变量之间的度量单位不同，对数据进行了标准化。分析数据之间的相关性，由输出结果可知，解释变量之间的相关系数都较大，从而解释变量之间就有可能存在多重共线性，于是本文使用 VIF 检验解释变量之间是否存在多重共线性，输出结果如表 4：

Table 4. Explanatory variables for VIF values

表 4. 解释变量的 VIF 值

变量	X1	X2	X3	X4	X5	X6	X7	X8	X9
VIF	66.9	319.2	4560.4	4371.7	414.7	446.1	132.2	405.6	325.0

注：该数据均只保留了一位小数。

由输出结果可知，每一个解释变量的 VIF 值都大于 10，此时使用最小二乘法建立的模型并不具有说服力，满足选择主成分回归方法的条件，所以选用主成分回归分析，又由于数据之间数量级相差较大，所以由相关矩阵求解主成分。

在输出结果中，第一个主成分的贡献率为 0.9159606，即第一个主成分含有九个变量中的信息高于百分之九十；前两个主成分的累计贡献率为 0.99011011，即前两个主成分包含的信息已经超过了百分之九十九，以此类推，九个主成分完全包含了九个变量的所有信息。为了保留原始数据信息，在选取主成分

时，本应该让累计贡献率尽可能的高，但在实际的应用中，我们更希望用尽可能少的主成分来概括原始数据的信息。在本次研究中，第一个主成分的贡献率已经高于 0.9，因此本次研究只需选取一个主成分即可。且第一个主成分与原影响变量之间的线性关系如表 5：

Table 5. Linear relationship of the first principal component to each of the original variables

表 5. 第一个主成分与各原始变量的线性关系

变量	X1	X2	X3	X4	X5	X6	X7	X8	X9
系数	0.279	-0.322	-0.343	-0.340	-0.346	-0.341	0.334	-0.344	-0.347

即第一个主成分的表达式为：

$$z_1 = 0.279X_1^* - 0.322X_2^* - 0.343X_3^* - 0.340X_4^* - 0.346X_5^* - 0.341X_6^* + 0.334X_7^* - 0.344X_8^* - 0.347X_9^*$$

式中 z_1 表示第一个主成分， X_i^* 为 X_i 的标准型。

3.4. 对中国人口出生率的主成分回归

3.4.1. 理论模型

由以上的输出结果可知，我们只选取了一个主成分，此时可将人口出生率(Y^*)作为被解释变量，将第一个主成分(z_1)作为解释变量，使用最小二乘回归的方法对其建立模型，本文根据以上可构建以下模型：

$$Y^* = b * z_1$$

模型中 Y^* 表示标准化以后的被解释变量人口出生率， b 为第一主成分与被解释变量之间的线性相关系数， z_1 为第一主成分。使用最小二乘回归法建立第一主成分与被解释变量之后，可将第一主成分与各原始影响变量的线性回归方程带入到模型 1 中，从而可得到人口出生率与各原始变量之间的线性模型如下：

$$Y^* = \beta_1 X_1^* + \beta_2 X_2^* + \beta_3 X_3^* + \beta_4 X_4^* + \beta_5 X_5^* + \beta_6 X_6^* + \beta_7 X_7^* + \beta_8 X_8^* + \beta_9 X_9^*$$

其中 β_i 表示各原始变量 X_i^* 与被解释变量之间的线性系数。

3.4.2. 实证分析

根据输出结果可知， Y^* 对第一个主成分做回归时，回归方程和回归系数均显著。且主成分回归方程为：

$$Y^* = 0.124044z_1$$

由于主成分是标准化后自变量的线性组合，如果想要得到 Y^* 关于标准化后的九个自变量的回归方程，只需分别将下面一个式子：

$$z_1 = 0.279X_1^* - 0.322X_2^* - 0.343X_3^* - 0.340X_4^* - 0.346X_5^* - 0.341X_6^* + 0.334X_7^* - 0.344X_8^* - 0.347X_9^*$$

代入上式可得到最终模型：

$$Y^* = 0.034607X_1^* - 0.03994X_2^* - 0.04255X_3^* - 0.04217X_4^* - 0.04292X_5^* - 0.0423X_6^* + 0.041429X_7^* - 0.04267X_8^* - 0.04304X_9^*$$

由最终模型可知,每一影响变量的系数绝对值均小于 0.1,且每一变量的系数绝对值大小相差并不大,都在 0.03 至 0.05 之间。输出的 R^2 反映了最终输出方程的拟合度,其表示的由回归方程引起的总离差平方和变动的占比, R^2 越接近于 1,可认为该方程拟合得越好。而该模型所输出的模型的 R^2 仅有 0.1335,所以可认为使用主成分回归拟合模型的效果并不理想。经查阅资料发现,在有些情况下,解释变量虽然对预测被解释变量有所帮助,但当该解释变量被包括在方程以后,被解释变量的不可解释的变差依然存在,从而导致 R^2 偏低。因此本文使用逐步回归的方法继续建立模型,观察当通过逐步回归选择剔除某些变量后方方程的拟合度是否会有所提高,并且可从中得到影响出生率的显著因素以及回归方程。

4. 基于逐步回归分析探索影响出生率的主要因素

4.1. 数据分析方法

由前述相关性知,数据之间存在多重共线性。因此本文为了消除解释变量之间的多重共线性并且得到相对最优的模型选择逐步回归的统计方法对数据进行分析(肖雪梦, 2015) [19]。逐步回归的基本思想是可以引进来也可以踢出去,其判断准则是 AIC (最小信息量准则),输出的模型在整体上是最优的。

4.2. 逐步回归分析结果

对数据进行逐步回归分析后输出如表 6,由表 6 可得,在使用逐步回归分析消除变量之间的多重共线性后, X_2 、 X_4 对出生率有高度显著影响, X_8 和 X_9 具有显著影响,剩余变量对出生率的影响并不显著。

Table 6. Stepwise regression output results

表 6. 逐步回归输出结果

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	5.54E-16	1.16E-01	0	1	
X_2	-2.232E+00	7.47E-01	-2.987	0.00921	**
X_4	3.67E+00	1.11E+00	3.299	0.00487	**
X_8	-3.834E+00	1.65E+00	-2.329	0.03424	*
X_9	1.925E+00	9.95E-01	1.934	0.07217	.

由输出的结果可知,模型的 R^2 为 0.7885,拟合效果较主成分回归好。在消除变量之间的多重共线性以后影响 Y 的显著影响因素为 X_2 、 X_4 、 X_8 、 X_9 ,其中 X_2 (老年抚养比)和 X_8 (商品房平均销售价格)与 Y (出生率)为负相关,即当 X_2 、 X_8 增加时, Y 减少; X_4 (居民消费水平), X_9 (参加生育保险人数)与 Y 成正相关的关系,当 X_4 、 X_9 增加时, Y 也随着增加。并由表 6 得出如下回归方程:

$$Y^* = -2.232X_2^* + 3.670X_4^* - 3.834X_8^* + 1.925X_9^*$$

综合以上研究,我们可知,使用主成分回归建立出来的模型,其拟合度较差。当使用逐步回归剔除 X_1 、 X_3 、 X_5 、 X_6 、 X_7 后建立的模型拟合度明显变好,从而可得将 X_1 、 X_3 、 X_5 、 X_6 、 X_7 加入方程后会影响 Y 的不可解释变差,进而导致其拟合度差。由逐步回归消除变量之间的多重共线性之后输出的结果可知,老年抚养比、居民消费水平、商品房平均销售价格及参加生育保险人数均对出生率具有显著的影响。其中,最显著的是老年抚养比和居民消费水平。

分析最终输出模型可知,老年抚养比与商品房平均销售价格均对出生率产生负面影响,影响系数分别为 -2.232 和 -3.834。由老年抚养比的计算公式可知,老年抚养比的升高也就意味着每一个适龄劳动力需要赡养的老人增多,直接加大了适龄劳动人口的经济压力,适龄劳动人口为了适当减轻经济压力,必

然会选择少生少育,从而直接导致了出生率的下降,即老年抚养对出生率具有负面影响。商品房平均销售价格对出生率有负面影响的结论与胡佩等人(2020) [5]的研究结论相匹配。房价与人民生活息息相关,随着房价的升高会推迟初次生育时间,初生时间推迟,随之而来的则是大龄产妇的危害,从而进一步导致生育率下降,即随着房价的上涨生育率会有所下降。居民消费水平、参加生育保险人数与出生率呈正相关的关系,其影响系数分别为 3.67、1.925。居民消费水平反映了劳务的质量和数量,由庄亚儿,姜玉等人(2021) [3]的研究可知,养育成本过高是导致育龄人群“不敢生”的一个重要因素,家庭中每出生一个婴儿,其家庭中消耗在婴幼儿产品中的金额数量必然会增加,进而导致整个家庭的消费总额增加,也就使得居民消费水平提高,这就是居民消费水平与人口出生率呈正相关的原因。生育保险是国家或社会对生育的职工给予必要的经济补偿和医疗保健的社会保险制度,所以参加生育保险的人数增加,则说明更多的职工倾向于生育孩子,生育率则随之升高。

5. 结论

由研究结论的主要影响因素可知,国家难以通过调整对出生率产生负面影响的老年抚养比来提高出生率,目前实施的二胎政策也远远没有达到预期的效果,虽然其可以对人口结构起着一定的优化作用,但其终究还是改变不了老年抚养比逐年上升的现实情况。因此国家只有不断优化生育政策并实施一些配套政策才能保持人口出生率不再下降或者促进出生率上升,针对现如今人口老龄化以及老年抚养比逐年增加的问题,国家应该更加完善社会保障制度,特别是对老年人的保障,应该增加养老保险的保障的范围,使得更多地区的老人能享受养老保险的保障,从而在一定程度上减轻适龄生育人群的经济压力。或者鼓励退休老年人再学习,鼓励其多参加实践活动,使得劳动力资源充分利用。对房价问题,国家应制定严格政策,贯彻落实“房子是用来住的,不是用来炒的”的提议,严格控制房价,让适龄人群也能有房住,从而促进生育率的提高。针对适龄人群“不敢生”的问题,国家可以给予一定的养育经济支持,适当增加公立学校的数量,并提高公立学校教学质量,提高老师的综合素质,从而让家长更放心将孩子送往公立学校学习,而不是为了让孩子“不输在起跑线上”花高额的学费将孩子送往私立学校,如此可以降低孩子的教育成本以及家庭抚养孩子的经济压力。并给予生育女性更多的工作以及生育保障,尽量降低其因为生育孩子造成的损失。

参考文献

- [1] 董璐. 人口老龄化对我国经济发展的影响研究[D]: [硕士学位论文]. 长沙: 湖南大学, 2019
- [2] 陈卫. “发展-计划生育-生育率”的动态关系: 中国省级数据再考察[J]. 人口研究, 2005(1): 2-10.
- [3] 庄亚儿, 姜玉, 李伯华. 全面两孩政策背景下中国妇女生育意愿及其影响因素——基于 2017 年全国生育状况抽样调查[J]. 人口研究, 2021, 45(1): 68-81.
- [4] 刘金菊. 中国城镇女性的生育代价有多大? [J]. 人口研究, 2020, 44(2): 33-43.
- [5] 胡佩, 王洪卫. 住房价格与生育推迟——来自 CGSS 微观数据的证据[J]. 财经研究, 2020, 46(4): 79-93.
- [6] 杨龙见, 陈建伟, 徐琰超. 财政教育支出降低了人口出生率? [J]. 经济评论, 2013(3): 48-55.
- [7] 李松臣, 张世英. 基于逐步回归法的人口出生率影响因素分析[J]. 统计与决策, 2008(4): 7-9.
- [8] 杨慧诗, 孟晗. 关于人口出生率影响因素的逐步回归分析[J]. 现代商业, 2010(5): 178+177.
- [9] 李建伟. 我国人口出生率的影响因素及其发展趋势[J]. 发展研究, 2014(9): 71-78.
- [10] 陈阔. 人口红利趋缓下人口出生率影响因素实证分析[J]. 品牌, 2015(12): 76.
- [11] 刘丽萍. 基于 Lasso 回归法的人口出生率影响因素分析[J]. 牡丹江师范学院学报(自然科学版), 2018(2): 1-3.
- [12] 玛依拉·吐尔逊, 常轩. 我国人口出生率及其影响因素研究——基于动态 GMM 模型和门限模型的实证分析[J]. 价格理论与实践, 2019(11): 53-56.

- [13] 田青. 基于 OMP 算法的人口出生率影响因素分析[J]. 广西质量监督导报, 2020(2): 74-76.
- [14] 孙鑫鑫, 高燕, 袁汐, 王增相. 我国人口出生率的分析与预测[J]. 山东师范大学学报(自然科学版), 2020, 35(3): 256-264.
- [15] 张祎, 朱家明. 基于 ARIMA 和二次指数平滑模型的人口组合预测[J]. 牡丹江师范学院学报(自然科学版), 2017(1): 9-12.
- [16] 张现苓, 翟振武, 陶涛. 中国人口负增长: 现状、未来与特征[J]. 人口研究, 2020, 44(3): 3-20.
- [17] 刘岩. 我国人口年龄结构变动对经济增长的影响研究[D]: [博士学位论文]. 长春: 吉林大学, 2020.
- [18] 任雪松, 于秀林. 多元统计分析[M]. 北京: 中国统计出版社, 2010: 393.
- [19] 肖雪梦, 张应应. 三种回归方法在消除多重共线性及预测结果的比较[J]. 统计与决策, 2015(24): 75-78.