

机器学习视角下出生率与地区经济发展研究

黄家祥

【摘要】使用LassoCV方法和聚类分析机器学习模型，将我国31个省市根据出生率和地区经济发展水平聚分为五类。研究发现，地区经济发展水平对出生率的影响是动态的，呈现出“抑制—促进—再抑制”的特征，出生率则呈现出“下降—上升—再下降”的趋势；人均地区生产总值、人均可支配收入及商品房平均销售价格是影响出生率的共同关键因素。

【关键词】出生率；经济发展；LassoCV；机器学习；聚类分析

一、问题提出

为抑制人口过快增长，我国于1982年将计划生育政策纳入基本国策。然而，随着改革开放的深化，在经济快速增长的同时，近年来我国反陷入低生育陷阱。尽管我国相继提出“单独二孩”和“全面二孩”的鼓励政策，但收效甚微。长期来看，较低的出生率不仅会影响经济的健康持续发展，更可能导致人口结构的失衡。

针对出生率下降的原因，马克思早在《资本论》中阐述过，随着科技的不断进步，资本有机构成迅速提高，致使劳动需求减少和智力投资增加，迫使人们不婚不育或迟婚少育。Schultz（1973）从成本的视角出发，认为父母工资率的上升会变相增加养育子女的时间成本和机会成本，从而抑制父母对子女数量的需求。Easterlin 与 Crimmins（1985）认为社会经济发展使人们的生育理念发生转变，就业竞争及生活不稳定导致人们推迟生育年龄。都阳（2005）提出人口转变是社会政策和经济发展共同作用的结果。杨华磊等（2018）指出城镇化的推进会轻微降低生育。严成樑（2018）通过跨期叠代模型发现，子女抚养成本是制约出生率提高的重要因素。关于出生率对社会经济发展的影响。Hussain（2002）指出，抚养人口的比例和构成、劳动年龄人口结构等都会对经济发展产生影响。都阳（2004）发现，人口因素对经济增长的影响是动态的。曹阳等（2018）、高冉

等（2018）通过三期世代交叠模型发现，出生率、少儿老年抚养比与国民储蓄呈显著的负相关。已有研究丰富了对出生率和经济发展的认知，但仍然存在一些研究空白，如出生率不断下降的关键因素，出生率与经济发展相互作用的机理等。基于此，本文拟针对上述问题做出补充和探究。

二、聚类分析及特征变量识别

尽管影响出生率的因素错综复杂，但大致可归结为宏观和微观两方面。宏观层面主要包括国家生育政策、税费制度设计及社会经济整体发展水平等；而微观层面则主要包括居民受教育水平、可支配收入状况等。基于客观性、系统性及可操作性原则，本文选取宏观经济发展水平、文化教育资源和人口数量结构作为一级指标，选取人均地区生产总值、人均可支配收入、第一产业占比、居民消费价格指数、商品房平均销售价格、人均教育经费、中小学总和生师比、年末人口数、少年儿童抚养比、老年人口抚养比、男女性别比、大专及以上学历人数占比作为二级指标，指标数据均来源于中国统计年鉴，极少缺失数据均采用线性插值法获得。

机器学习是指通过对训练数据的学习，提升机器的性能，从而能够从无序的数据中提取有用的信息，利用习得的规则对新数据进行预测和分析的过程。聚类分析属于机器学习模型中的无监督学习。为深入研究中国当前出生率与地区经济发展现状，本文选取2017年的人均地区生产总值和出生率数据，采用机器学习模型中的聚类算法对31个省市进行聚类分析，并根据肘算法将31个省市聚分为五类，结果见表1。由表1聚类中心点研究发现，出生率随着人均地区生产总值的增长呈现出“先下降—后上升—再下降”的趋势。本文选取距离各聚类中心点较近的安徽、辽宁、陕西、福建、上海共5个省市作为核心研究对象。

表1基于人均地区生产总值—出生率的聚类结果

聚类名称及类别	聚类中心点	聚类省市
低人均生产总值 ——高出生率（Ⅰ）	（4.25，14.07）	河北、安徽、江西、河南、湖南、广西、海南、贵州、云南、西藏、甘肃、青海、宁夏、新疆
较低人均生产总值 ——低出生率（Ⅱ）	（5.01，6.49）	辽宁、吉林、黑龙江
中人均生产总值 ——较高出生率（Ⅲ）	（5.52，11.11）	山西、内蒙古、湖北、重庆、四川、陕西

较高人均生产总值 ——高出生率 (IV)	(8.21, 14.54)	浙江、福建、山东、广东
高人均生产总值 ——较低出生率 (V)	(12.04, 8.63)	北京、天津、上海、江苏

Lasso是一种利用压缩估计的思想,将参数估计与变量选择同时进行的一种正则化的方法。LassoCV的损失函数及优化方法与Lasso相同,区别在于验证方法。LassoCV对超参数使用交叉验证,寻找最合适的。本文采用LassoCV方法识别影响出生率的关键变量,其结果见表2。

表2基于LassoCV方法识别的影响出生率的关键因素

变量	安徽	辽宁	陕西	福建	上海
人均地区生产总值	0.00002	-0.00004	-0.00003	0.00003	0.00015
人均可支配收入	-0.00014	0.00024	0.00001	0.00016	0.00014
第一产业占比					
居民消费价格指数				-0.10531	0.66224
商品房平均销售价格	0.00154	-0.00114	0.00047	-0.00102	-0.00031
人均教育经费	-0.00101	0.00068		0.00565	0.00282
中小学总和师生比					0.91055
年末人口数	0.00059		0.00785	-0.00937	0.00914
少年儿童抚养比					-3.14823
老年人口抚养比					-0.25278
男女性别比				-0.04051	-0.70645
大专及以上学历人数占比				-0.07171	0.26743

三、出生率与地区经济发展互动关系及作用机理

出生率与地区经济发展相互作用,且在不同发展阶段呈现差异性和动态化的特征。首先,地区经济发展水平对出生率的影响是动态变化的,呈现出“抑制—促进—再抑制”的动态特征。表1显示,从第Ⅰ类向第Ⅴ类演进时,出生率呈现出“下降—上升—再下降”的趋势。当地区经济发展水平由较低向中等过渡时,地区经济的发展第一次压缩了人们的生育意愿,经济发展激励人们花费更多的资源去改善生活质量,因而花费在生育上的资源就会减少,出现生育资源被挤占现象。当经济进一步发展,居民反过来将视野集中到生育和培养后代上来,出生率会随着地区经济的发展而不断提升。当经济发展水平较高时,居民生活成本较重,居民人均可支配收入已无法与地区经济发展水平相匹配,致使地区经济发展第二次压缩了人们的生育意愿并再次挤占生育资源,使生育率下降。其次,人均地区生产总值、人均可支配收入以及商品房平均销售价格是影响出生率的关键因素,而人均地区生产总值增速与人均可支配收入增速的不相匹配是导致出生率下降的重要原因。最后,地区经济发展到一定水平后,人们对孩童的抚养教育

更加重视,少年儿童抚养比、中小学总和师生比等成为影响出生率的重要因素。

(重庆第二师范学院 数学与信息工程学院, 重庆 400065)

参考文献:

- [1]曾德国.战后发达国家人口自然增长率长期下降的原因及其利弊[J].武汉大学学报(社会科学版),1982(1).
- [2]Schultz.New Economic Approaches to Fertility[M]. Chicago:University of Chicago Press.
- [3]Richard A. Easterlin, Eileen M. Crimmins.The Fertility Revolution:A Supply-Demand Analysis[M]. Chicago: University of Chicago Press,1985.
- [4]都阳.中国低生育率水平的形成及其对长期经济增长的影响[J].世界经济,2005.
- [5]杨华磊,吴义根,张冰鑫.城镇化、外部性与生育水平[J].人口与发展,2018.
- [6]严成樑.老年照料、人口出生率与社会福利[J].经济研究,2018(4).
- [7]Hussain.Demographic Transition in China and Its Implications[J].World Development,2002,30(10): 1823-1834.
- [8]都阳.人口转变的经济效应及其对中国经济增长持续性的影响[J].中国人口科学,2004(5).
- [9]曹阳,徐升.抚养负担与储蓄关系再检验:基于年轻人储蓄视角的研究[J].人口与发展,2018(5).
- [10]高冉,蔡雯霞,张冲.人口老龄化、出生率与储蓄率的关系研究—兼论对中国“二孩政策”的启示[J].武汉金融,2018(1).
- [11]阮敬.Python数据分析基础[M].北京,中国统计出版社,2017.
- [12]赵志勇.Python机器学习算法[M].北京,电子工业出版社,2017.
- [13]谭常春,谭景宝,朱华亮.城市火灾次数与气象因素的Adaptive—Lasso分析[J].应用数学与计算数学学报,2013(9).

作者简介:黄家祥,(1984—)男,汉,讲师,研究生学历,经济学硕士学位,工作单位:重庆第二师范学院 数学与信息工程学院。