

中国人口转变地域差异的聚类分析

李建民 原 新 王金营

【提要】 本文通过对各省份的人口出生率、死亡率和自然增长率的聚类分析,从定量与定性两方面论述,把中国人口转变分为都市、东部、中部、西部和西藏五种模式。

【作者】 李建民 南开大学人口与发展研究所所长、教授;原新 南开大学人口与发展研究所,副教授;王金营 河北大学人口研究所,助理研究员。

20 世纪中国的人口转变呈现出地域差异性和非同步性。本文旨在对各省 1949~1999 年人口出生率、死亡率和自然增长率变动趋势进行综合比较分析的基础上,采用聚类分析方法对人口转变的地域差异及不平衡性作出定量和定性判断,归纳出中国人口转变的不同区域模式。

一、数据来源

中国不同来源的统计资料对分省的出生率、死亡率和自然增长率的记录差异颇大。通过比较,我们选择了 80 年代出版的《中国人口》丛书和 90 年代出版的《跨世纪的中国人口》系列丛书、《中国人口续篇:80 年代中国人口变动分析》作为数据的主要来源,90 年代以后的数据主要引自国家统计局公布的历年人口的统计资料。

二、研究方法

在比较分析的基础上,选择人口出生率、死亡率和自然增长率 3 个指标作为聚类变量分别对各省进行聚类分析,结合实际统计资料,定量和定性相结合揭示中国人口转变区域发展模式及特征。

(一) 聚类分析的原则

首先,确定变量。其原则是:(1)聚类分析目标密切相关的变量;(2)反映被分类对象特征的变量;(3)不同研究对象的值具有显著差异的变量。其次,确定分类数。遵循的准则是:(1)任何类都必须在邻近各类中是突出的,即各类中心值之间的距离必须足够大;(2)各类所包含的元素不能过分地多;(3)分类的数目应该符合使用的目的;(4)若采用几种不同的聚类方法,则在各自的聚类图上应发现相同的类。经过比较分析,我们按 80 年代中国大陆地区的行政区划,将当时的 28 个省、自治区、直辖市^①分成四类。

(二) 聚类方法与步骤

研究中采用简单中心聚类方法,分别对出生率、死亡率和自然增长率进行聚类分析,根据分析的目的和需求取舍聚类变量。

1. 人口出生率水平的聚类方法。首先,按曲线形态聚类。第一步将各省(市、区)1949~1999 年出生率的变动趋势用曲线描述,估算出各省(市、区)出生率变动拟合曲线的指数,利用估算得到的参数作为聚类变量进行聚类。在曲线拟合时,发现 S 曲线和 Logtis 曲线的拟合优度均较高且二者差异极小,因此我们只选取 S 曲线作为各省(市、区)出生率数据的拟合曲线形态。S 曲线的函数形式为:

^① 本研究未包含港、澳、台地区。西藏自治区因数据系列不完整,未纳入聚类分析中。

$Y_t = e^{(b_0 + \frac{b_1}{t})}$ 。第二步利用各省(市、区)出生率历史数据,估计各自 S 曲线函数中的 b_0 、 b_1 值和曲线的拟合优度 R^2 ,用参数 b_0 、 b_1 值刻画各省(市、区)出生率变动曲线的特征,用 R^2 反映估计曲线与实际曲线的拟合程度。第三步将各省(市、区)的 R^2 和参数 b_1 作为聚类变量进行聚类,目的是区分各省(市、区)之间出生率变动形态的异同。其次,按分时期的截面水平聚类。在 1949~1999 年期间选取 4 个时期各省(市、区)的出生率水平作为聚类变量,目的是反映各省(市、区)在该时期出生率变动的类别差异。

2. 人口死亡率水平的聚类方法。与出生率水平的聚类方法相同,只是以死亡率作为聚类对象。

3. 人口自然增长率水平的聚类方法。由于各省(市、区)的自然增长率曲线跌宕起伏,形态迥异,进行拟合时几乎找不到一条有代表性的曲线来刻画各省(市、区)自然增长率的变动趋势。所以,在实际分析中为了消除自然增长率曲线振幅不稳定的影响,选取了 1955 年以后每隔 5 年以及 1999 年各省(市、区)的自然增长率作为聚类变量,采用简单中心聚类方法进行聚类。

4. 综合分析方法。将上述分析得到的聚类结果与各省(市、区)实际人口出生率、死亡率和自然增长率进行综合分析比较,确定中国人口转变的区域模式。

三、聚类分析的结果

(一) 各省(市、区)人口出生率聚类结果

首先利用 SPSS 软件对各省(市、区)出生率的历史数据进行 S 曲线拟合,得到参数 R^2 、 b_0 、 b_1 (见表 1);然后对 R^2 、 b_0 、 b_1 三个变量进行相关分析,发现 b_0 、 b_1 两个参数高度相关,因此在选择聚类变量时只选择 b_1 与 R^2 ,对各省(市、区)进行简单中心值聚类(见表 2);最后将聚类结果与各省(市、区)出生率实际曲线进行比较,并根据 S 曲线的性质得到各类地区出生率曲线变动的特征:

表 1 1949~1999 年中国各地区人口出生率的 S 曲线参数

地区	R^2	b_0	b_1	地区	R^2	b_0	b_1	地区	R^2	b_0	b_1
北京	0.787	-63.452	130 930	江苏	0.676	-47.478	99 699	广西	0.547	-30.751	67 196
天津	0.788	-47.125	98 838	浙江	0.856	-47.563	99 940	四川	0.326	-31.100	67 503
河北	0.637	-30.234	65 675	安徽	0.18	-19.535	44 755	贵州	0.369	-23.032	52 147
山西	0.593	-27.625	60 740	福建	0.643	-36.666	78 718	云南	0.448	-21.910	49 815
内蒙古	0.794	-41.751	88 892	江西	0.596	-33.522	72 620	陕西	0.710	-33.442	72 252
辽宁	0.742	-48.458	101 625	山东	0.715	-44.721	94 345	甘肃	0.380	-27.896	61 360
吉林	0.811	-51.525	107 944	河南	0.495	-31.794	69 017	青海	0.238	-23.971	53 845
黑龙江	0.791	-56.845	118 449	湖北	0.591	-36.023	77 285	宁夏	0.479	-32.415	70 639
上海	0.782	-74.005	151 456	湖南	0.619	-40.109	85 406	新疆	0.516	-20.656	47 283
				广东	0.711	-30.289	66 178				

I 类地区的出生率曲线与 S 曲线的拟合程度高,曲线弧度大,表明出生率曲线波动幅度小、下降的速度快;II 类地区的出生率曲线与 S 曲线的拟合程度比较高,曲线弧度略小,表明出生率曲线波动幅度较小,下降的速度略慢于 I 类地区;III 类地区的出生率曲线与 S 曲线的拟合程度较低,曲线弧度较小,表明出生率曲线波动幅度较大,下降的速度较慢;IV 类地区的出生率曲线与 S 曲线的拟合程度低,曲线弧度几乎为 0,表明出生率曲线波动幅度大,总体下降的速度慢。

事实上,各省(市、区)出生率转变都是在 1949 年以后出现的,虽然彼此间在转变启动时间、转变速度和完成时间上差异巨大,但在转变形态上,大多数地区具有相同性。

1. 出生率转变启动时间。基本上是自东向西各省递次进入出生率转变过程,依次为大城市和东部经济条件好的省(市)最先开始(上海、北京、天津、辽宁、吉林、黑龙江、江苏、浙江始于 50 年代

表2 1949~1999年中国各地区人口出生率S曲线聚类结果

类别	地 区	中心值	
		R^2	b_1
I	北京、上海	0.785	1 141 193
II	天津、内蒙古、辽宁、吉林、黑龙江、江苏、浙江、山东	0.772	101 216
III	河北、山西、福建、江西、河南、湖北、湖南、广东、广西、四川、陕西、甘肃、宁夏	0.564	70 356
IV	安徽、贵州、云南、青海、新疆	0.350	49 569

中后期);然后是东部其他省份(河北、山西、安徽、福建、山东、湖北、湖南始于60年代中后期);接着是中部省(区)(内蒙古、江西、河南、广西、四川、陕西始于70年代初);西部省(区)最迟(甘肃、青海、宁夏、云南、贵州、新疆始于70年代中后期);最早与最迟两类地区至少相差20年;西藏自治区出生率的转变尚未真正开始。

2. 出生率转变完成时间。考虑到国际上普遍认可的低自然增长率上限标准是10‰,以及中国死亡率基本稳定在6‰左右的事实,我们将16‰作为判定低出生率水平的上限标准,发现各地区出生率转变完成的时间差别非常大。1966年上海市率先完成转变;北京、天津和江苏3省、市随后在70年代完成;浙江和辽宁两省在80年代完成转变;90年代前期完成转变的有山东、吉林、河北、黑龙江、湖南、河南、福建和陕西等省;90年代后期完成转变的有安徽、内蒙古、湖北、四川、广西、山西和甘肃省;西藏、贵州、青海、云南、新疆、宁夏、海南和江西省(区)20世纪末仍处于转变之中,其中西部各省(区)1999年的出生率都在17‰以上,有3个省(区)甚至超过20‰。

3. 出生率转变经历的时间。选取出生率峰值年和出生率稳定在16‰及以下的起始年两个时点,其间隔时间就是出生率转变所经历的时间^①,以此作为比较各省(市、区)出生率转变过程时间长短的参照系。大城市用了不足20年,东部和中部的绝大部分省(区)经历了30~40年,西部的多数地区至今尚在转变中,至少要经历50年以上。已经完成转变的地区中,上海市出生率从52.6‰降到14.5‰,仅用了11年,而广东省出生率从41.94‰降到15.32‰经历了48年,是上海市的4.3倍。

(二) 各省(市、区)死亡率水平聚类结果

利用SPSS软件对各省(市、区)死亡率历史数据进行S曲线拟合,得到参数 R^2 、 b_0 、 b_1 (表见3),并根据所得参数对各省(市、区)死亡率进行聚类(见表4),结果显示:

I类地区人口的死亡率变动曲线已经呈现反“J”形状,表明由于人口年龄结构的变动,人口老龄化程度提高,近几年的死亡率水平已经开始提高;II类地区的特点接近第一类地区,但死亡率水平刚开始出现上升迹象;III类地区的死亡率仍处于下降通道,但继续降低的空间十分有限;IV类地区的死亡率仍在下降,而且下降空间相对较大。

经典人口转变理论认为,人口转变过程以死亡率的转变为发端,该论断在中国人口转变实践中得到证明。虽然缺乏1949年以前的人口死亡率资料,但是可以肯定,死亡率的转变是在新中国成立以后才真正开始的。1949年死亡率高达20‰^②,1955年降到了12.28‰,70年代中后期又降低到7‰以下,在随后的20多年中基本上保持在6‰~7‰。但各地区死亡率所展现出的变化却更加丰富。

1. 死亡率转变启动时间。由于缺乏1949年以前各省(市、区)死亡率资料,对各省(市、区)死亡率

① 有3种情况需要说明:(1)1963~1964年出现的生育高峰不考虑;(2)有些地区在一段时间内出生率曾经降低到16‰以下(不考虑“三年自然灾害”时间),但一般持续了4~5年后又提升到16‰以上,不能算转变完成;(3)还有一些省份的出生率直到1999年才降到16‰,虽然今后的稳定性还不能确认,我们仍然把1999年视为其实现低生育率的起始年份。

② 联合国估计值更高,1950~1955年死亡率为25‰,略低于同期最不发达国家27.8‰的水平。

表3 1949~1999年中国各地区人口死亡率S曲线拟合结果

地区	R ²	b ₀	b ₁	地区	R ²	b ₀	b ₁	地区	R ²	b ₀	b ₁
北京	0.640	-21.278	45 762.9	江苏	0.559	-30.449	64 163.8	广西	0.572	-28.778	60 611.4
天津	0.435	-16.607	36 596.4	浙江	0.477	-20.678	44 734.3	四川	0.443	-45.954	95 260.4
河北	0.587	-31.630	66 492.6	安徽	0.419	-29.622	62 323.8	贵州	0.561	-44.893	93 322.1
山西	0.756	-37.775	78 759.6	福建	0.496	-21.246	45 742.4	云南	0.598	-35.205	74 055.7
内蒙古	0.445	-23.135	49 541.2	江西	0.642	-31.673	66 721.8	西藏	0.628	-32.915	69 176.8
辽宁	0.312	-22.601	48 325.8	山东	0.638	-39.185	81 593.0	陕西	0.534	-43.764	90 596.4
吉林	0.638	-29.578	62 312.7	河南	0.536	-37.004	77 195.0	甘肃	0.360	-34.252	71 671.1
黑龙江	0.628	-35.558	73 918.4	湖北	0.609	-28.743	60 950.6	青海	0.624	-45.422	93 545.9
上海	0.042	5.449	-71 16.7	湖南	0.624	-36.739	76 901.6	宁夏	0.626	-36.268	75 941.2
				广东	0.511	-38.805	80 757.7	新疆	0.516	-20.656	47 283.0

表4 1949~1999年中国各地区人口死亡率的S曲线聚类结果

类别	地 区	中心值	
		R ²	b ₁
I	上海	0.042	5.449
II	北京、天津、内蒙古、辽宁、浙江、福建	0.468	-20.924
III	河北、山西、安徽、云南、江西、河南、湖北、湖南、广西、青海、新疆、广东、陕西、山东、吉林、黑龙江、江苏	0.584	-33.775
IV	四川、宁夏、贵州、甘肃	0.541	-45.008

转变的启动时间做出精确判断是困难的,但根据50年代初各省(市、区)的死亡率水平,大致可分为两类:一类是在新中国成立以前就已经进入转变过程的地区,如1953年人口普查,死亡率低于10‰的地区有上海、辽宁、四川^①;低于12‰的地区有北京、天津、河北、吉林、黑龙江、山东、甘肃。另一类是在新中国成立以后才开始死亡率转变过程的地区,如宁夏、新疆、河南、云南、湖南、贵州等;虽然缺失西藏、湖北、青海、江西50年代初的死亡率资料,但根据经验推断,这些地区的死亡率在当时应该处于高水平。

2. 死亡率转变完成时间。如果把死亡率相对稳定在6‰左右作为死亡率转变完成的判断标准,各省(市、区)死亡率转变完成时间的分布规律为:(1)80年代以前已有2/3的省(市、区)完成了死亡率转变;(2)西部地区死亡率转变明显落后于东部;(3)1949年以后开始死亡率转变的地区所经历的转变时间从15年到30年不等;(4)只有西藏自治区在90年代末尚未完成死亡率转变。

(三) 人口自然增长率聚类结果

选择各省(市、区)1955年以来每隔5年及1999年的自然增长率作为聚类变量,聚类结果如表5。

I类地区的起始人口自然增长率水平高,但下降速度最快,达到的水平最低;II类地区的起始人口自然增长率较高,下降速度较快,达到的水平较低;III、IV类地区的起始人口自然增长率较低,先经历一个上升时期,然后再下降,目前仍在继续降低。

自然增长率是由出生率和死亡率派生的指标,是出生率和死亡率两种相对立的力量作用的结果。自然增长率的变化反差最强烈的是上海市和西藏自治区。上海市早在1993年人口(户籍)就进入了负自然增长,1999年西藏自治区的人口自然增长率仍高达15.80‰;但是更重要的现象是越来越多的省(市、区)的人口自然增长率开始小于10‰,1975年只有京、津、沪3个直辖市,1999年增加到23个省(区),其中有13个省(区)的自然增长率已经接近或低于5‰。

根据预测,未来各省(区)实现零增长的时间也相差甚远,四类地区经过零人口增长点进入负增

① 统计资料显示安徽省1949年的死亡率仅为6.97‰,其后的4年也都在7‰左右,我们认为这些数据不可信。

长的时间递次出现,总的时间跨度为40~50年,表现为每10年一个台阶。上海已经实现负增长,北京、天津将在2015~2020年经过零人口增长点;Ⅱ类地区为2030年前后;Ⅲ类地区推迟到2040年左右;Ⅳ类地区的部分省份将在2045年以后达到零增长,而新疆、西藏和云南均为少数民族地区,因地制宜的少数民族生育政策相当宽松,所以在

2050年以前还不会出现零人口增长,人口增长的势头将持续到21世纪后半期(李建民等,2000)。

表5 按地区自然增长率聚类结果

类型	地区
I	北京、天津、上海
II	河北、山西、内蒙古、辽宁、浙江、吉林、黑龙江、江苏、安徽、陕西、山东
III	福建、江西、湖北、湖南、广东、广西、新疆、云南、宁夏
IV	河南、四川、贵州、青海、甘肃

四、中国人口转变区域发展模式

根据上述分析,可以把中国的人口转变区域发展模式归纳为5种:

1. 都市模式。包括上海、北京和天津市,以上海市最为典型。人口死亡率在1949年以前就开始转变,新中国成立以后进一步迅速下降,并很快降低水平,90年代以来受人口年龄结构的影响已经显示出回升的迹象,曲线形态呈反“J”形状;出生率经历了抬高、然后在短期内迅速下降的过程,而且在下降过程中波动幅度小、下降速度快、稳定性强,最突出的特征是人口转变迅速、历时短暂。

2. 东部模式。包括浙江、江苏、辽宁、吉林、黑龙江、河北、山东和福建8个省,以辽宁和浙江为典型。死亡率转变开始于1949年以前,50年代以后的变动趋势与都市模式相近,但下降速度相对较慢,90年代末开始出现回升趋势;出生率变动方式与都市模式大致相同,而下降的速度明显慢于都市模式,转变时间相对延长,基本上在80年代中后期完成了生育率转变。

3. 中部模式。包括湖北、湖南、四川、陕西、河南、山西、江西、广东、广西、安徽、内蒙古和甘肃12个省(区),以湖南为典型。死亡率转变的初期在较高水平上维持时间相对较长,60年代中期开始形成迅速下降的势态,90年代末仍然处于下降状态,但继续下降的空间十分有限;出生率首先经历了一个由低到高的增加过程,高生育率持续的时间相对较长,出生率峰值出现在60年代中期,然后开始由高向低的转变,直到90年代中后期才完成生育率转变过程。

4. 西部模式。包括贵州、云南、青海、宁夏和新疆5个省(区)。死亡率变动形态与中部模式接近,但是处于高位死亡率时的水平更高,下降的速度慢于中部模式,90年代末仍有较大下降余地;出生率首先经历了由低到高的增长阶段,并在高位持续了更长时期,直到70年代中期才开始进入下降通道,但是下降速度慢、波动大、历时长,目前仍保持较高出生率水平。

5. 西藏模式。死亡率还处于转变过程之中;出生率刚呈现转变的迹象,而且具有与中国其他地区完全不同的特点,1965年(人口统计最早年份)西藏自治区出生率为14.13‰,然后进入一个持续上升的过程,1970达到25.26‰,但在随后的近20年里基本在25‰上下波动,直到90年代后期才出现明显的下降迹象,从1993年的26.68‰下降到1999年的23.20‰,在西藏自治区人口发展过程中,出生率并没有出现像中国其他地区那样的高峰。西藏的人口再生产是一个非常特殊的类型,即相对于其所处的人口转变阶段而言,具有“低出生、高死亡率、低增长”的特征。到目前为止,西藏的人口转变可以被视为是无政策干预的自发过程。

参考文献:

1. 国家统计局:《中国统计年鉴》(1990~1999年),中国统计出版社,1990~1999年。
2. 国家统计局人口与就业统计司:《中国人口统计年鉴》,(1990~1999年),中国统计出版社,1990~1999年。
3. 孙竞新主编:《跨世纪的中国人口》系列丛书(各省分册),中国统计出版社,1994~1996年。
4. 孙敬之主编:《中国人口丛书》(各省分册),中国财政经济出版社,1986~1991年。
5. 孙敬之主编:《中国人口续篇:80年代中国人口变动分析》,中国财政经济出版社,1996年。

(本文责任编辑:朱 萍)