

人口大流迁时期区域人口统计 指标的效度失真

梁秋生

【摘要】20世纪80年代以来,中国一直处于一个史无前例的人口大流迁时期。区域间大规模的人口流迁不仅对中国社会经济的发展和人口布局带来了巨大变化,而且还对传统的人口统计与分析方法带来一系列新的问题。文章认为,大规模的人口流迁不仅会造成一系列区域人口统计指标的效度失真,同时还会导致一系列与人口有关的社会经济统计指标的效度失真。

【关键词】人口流动 区域人口统计指标 效度失真

【作者】梁秋生 河北师范大学人口研究所,教授。

一、中国正处于一个史无前例的人口大流迁时期

自20世纪80年代以来,大量人口的流动与迁移已经成为中国最重要的人口现象之一。虽然中国近几次人口普查和全国人口抽样调查所界定的流动人口定义不同,但有学者估算,1982年全国流动人口规模仅为1100万,到1990年增加到2900万。根据2000年人口普查统计,当年全国流动人口总量为1.21亿,其中跨省流动人口总量为4242万人。2005年全国1%人口抽样调查显示,全国流动人口规模仍在继续扩大,当年全国流动人口数量已高达14735万人,其中,跨省流动人口4779万人。与第五次全国人口普查相比,跨省流动人口增加了537万人(国家统计局,2006)。近几年来,全国流动人口的数量还在继续增加,2008年全国流动人口达到2.01亿,2009年达到2.11亿。“十二五”期间,中国人口流动规模还将保持在2亿以上(国家人口和计生委,2010)。据此计算,1982~2009年全国流动人口平均每年以大约782万人的规模扩大。

除非户籍流动人口外,正规的户籍迁入人口在大城市和经济发达地区的常住人口中也占有相当大的比例。以高等院校在校生为例,1980年全国高等院校在校生仅为115.5万人,1990年发展到215.6万人,2000年增加到586.2万人,2008年猛增到2149.8万人。从2008年起,全国普通高校的年招生规模首次突破600万人,达到607.7万人(国家统计局,2009)。根据教育部和国家发展改革委员会下达的2010年全国普通高等教育招生计划,2010年全国普通高校研究生、本专科生的招生规模将达到710.4万人。以此推算,至2010年底,

中国普通高校在校生规模将超过2500万人。这说明,目前仅高等教育规模的超常发展,就使全国每年正规的户籍迁移人口超过700万人。此外,户籍待定人口、出国学习和工作的人口在中国城镇地区,特别是大中城市中也已达到相当的规模,他们也可以被看做某种形式的迁移人口或准迁移人口。毫无疑问,无论从规模还是从迁移形式及影响范围上看,中国目前正处于一个史无前例的人口大流迁时期。

二、在人口大流迁时期,区域人口统计指标的效度失真

庞大的流迁人口给中国社会、经济、文化及人口自身的发展等方面带来巨大的影响,在促进经济发展和社会进步的同时也引发了一系列的问题。在人口大流迁时期,区域人口统计指标和一系列与人口有关的社会经济指标的效度失真就是其中之一。

信度和效度是社会调查研究中的基本概念。所谓信度是指统计指标的可靠性、可信性,指测量结果(统计指标)的前后一致性或稳定性,而效度则是指统计指标(调查结果)所要说明问题的正确程度。笼统地说,信度就是指统计指标的准确性问题,而效度则是指统计指标能否正确反映统计对象的属性或数量特征的问题。任何社会经济统计指标(包括人口统计指标)都存在信度和效度的问题。在中国,有关人口统计指标信度问题的研究可谓硕果累累,但有关人口统计指标的效度问题,特别是在人口大流迁时期,区域人口统计指标的效度问题却没有引起同样的重视。

众所周知,人口流动与迁移的流向具有明显的规律性。就中国而言,中西部农村地区是流迁人口的主要来源地,而大中城市和东南沿海的经济发达地区则是中国流动人口的主要吸纳地。根据2000年全国人口普查统计,在当年跨省流动的4242万人中,有3161万人流入全国的城镇地区,占全部省际流动人口的74.52%,其中仅京、津、沪的外来流动人口就占全国的16.67%。2000年,北京市的外来人口为246.32万人,占总人口的18.15%;天津市的外来人口为73.50万人,占总人口的7.35%;上海市的外来人口为313.29万人,占总人口的18.72%。最新的统计数据显示,2009年北京市外来人口达到509.2万人,占总人口的29%;天津市外来人口达到265.99万人,占总人口的21.7%;上海市外来人口达到541.93万人,占总人口的28.21%。

此外,由于中国高等学校主要集中在大中城市,近10年来随着中国高校招生规模的不断扩大,城市中属于正规户籍迁移人口的高校在校生规模也在快速增加。仍以京、津、沪为例,2000年末,三市的高校在校生人数分别为35.16万、12.97万和26.49万,到2009年末,分别上升到78.60万、44.37万和61.63万,9年分别增长了123.55%、242.10%和132.65%。这说明,21世纪以来,京、津、沪等大城市仍然是外来流入人口^①最为集中和增加最快的地区。

外来流入人口的年龄结构、性别结构、社会构成及生活方式与流入地的“本土人口”具有极大的差异性,他们的人口行为后果(如婚姻、生育、死亡等)和众多的社会经济行为(如

^①为了叙述上的方便,本文将外来流入人口和高校在校生合称为外来流入人口或外来人口。

住房需求、抚养行为等)也与“本土人口”大不相同。从本质上看,外来流入人口不是传统意义上的稳定常住人口,而是一种动态变化的人口。基于中国现行的人口统计口径,外来流入人口的属性和特征使中国人口流入较为集中的地区,特别是大中城市的众多人口统计指标均存在不同程度的效度失真现象。大量的外来流入人口首先造成的是流入地区人口年龄和社会结构的显著变化。以年龄结构为例,由于大量外来人口的涌入,使中国城镇地区的人口年龄结构类型普遍年轻于农村地区和全国总人口的年龄结构类型。尽管 21 世纪以来中国流动人口的规模仍在快速增长,但由于缺乏分年龄的详尽人口统计数据,所以下面以 2000 年第五次全国人口普查数据来说明这一点(见表 1)。

表 1 2000 年全国、城、镇、乡及京、津、沪人口年龄构成及抚养比指标

地区	0~14 岁	15~64 岁	65 岁及以上	总抚养比	少儿抚养比	老年抚养比	%
全国	22.90	70.00	7.10	42.86	32.71	10.15	
城市	16.58	76.75	6.67	30.29	21.61	8.69	
镇	21.66	72.35	5.99	38.22	29.93	8.28	
乡村	25.52	66.98	7.50	49.30	38.10	11.20	
北京	13.59	77.99	8.42	28.22	17.42	10.80	
天津	16.76	74.82	8.41	33.65	22.40	11.24	
上海	12.26	76.28	11.46	31.09	16.07	15.02	

资料来源:国务院人口普查办公室、国家统计局人口和社会科技统计司:《中国 2000 年人口普查资料》,中国统计出版社,2002 年。

从表 1 可以看出,2000 年中国城镇地区的少儿人口系数和少儿抚养比均低于全国的平均水平,更低于乡村人口的平均水平。在中国现行的计划生育政策和人口统计口径下,这似乎并没有什么矛盾的地方。但就老年人口系数和老年抚养比也同样低于全国的平均水平而言,确是一个令人怀疑的问题。由于中国长期以来执行有差异的城乡生育政策,就常理而言,中国城镇人口的老龄化程度应明显高于全国的平均水平,更应该高于全国农村的平均水平,然而,统计结果却恰恰相反。显然,造成这种不合逻辑的统计结果是现行常住人口的统计口径所致,大量的外来劳动年龄人口必然导致城镇地区人口抚养比统计指标的普遍偏低。

当然,我们并不是怀疑这种统计结果的准确性,而是担心用这样的统计结果能否正确地说明问题的本质属性和特征。或者说,在利用这些指标对城镇地区的人口结构或相应的社会经济现象进行分析时,是否能够真正有效地说明问题。比如,用现行较低的老年抚养比指标能否真实地反映中国城镇人口现实和长远的抚养水平?这是值得深思的。

众所周知,大量的外来流入人口具有“两栖”的特征,他们虽然目前居住、劳动和生活在城镇地区,但他们当中许多人真正的被抚养对象却滞留在农村地区,形成了大量的留守儿童、留守妇女、留守老人等社会问题。无论从现实还是长远的角度看,这些外来流入人口的抚养功能并不能完全体现在城镇地区。从宏观统计的结果看,他们虽然降低了中国城镇地

区人口抚养比的统计水平,但从现实和长远的角度看,却起不到这样的作用。

从表1的统计数据来看,京、津、沪三大城市的少儿系数和少儿抚养比均低于全国的平均水平,而老年人口系数和老年抚养比都高于全国的平均水平,这看起来似乎符合现实的逻辑判断,其实并非如此。事实上,京、津、沪人口的年龄结构统计指标同样受外来人口的强烈影响,只是表面上与我们的基本判断不矛盾,实际上这种影响的力度比全国城镇地区更为强烈。从本质上说,像京、津、沪这样特大城市人口年龄结构的老化程度远比现实的统计结果严重,其实际的抚养系数也与现实的统计结果大不相同。

由于流入人口的年龄主要集中于育龄人口或劳动年龄人口,使流入地区,特别是大中城市地区的人口结构在统计上发生了巨大的变化,从而使众多受人口年龄结构影响的人口统计指标或多或少地存在效度失真的问题。以北京市为例,2000年北京市的人口粗死亡率仅为5.30‰,明显低于全国5.92‰的水平,而根据人口发展的一般规律,在人口出生率较低、人口平均预期寿命较长、社会现代化程度较高的地区,其粗死亡率指标应该明显高于生育率高、人口平均预期寿命较短、社会现代化程度较低的地区。可以推断,北京市的人口粗死亡率之所以低于全国的平均水平,与大量外来流入人口有直接的关系。由此判断,中国城镇地区现实的众多与人口结构有关的所谓“粗人口统计指标”,如出生率、死亡率、一般生育率、结婚率等,在功能上都或多或少地存在效度失真的问题,即这些指标不能正确地反映和说明上述各种现象本质的数量特征。

不仅如此,受大量外来流入人口的影响,一些传统上认为不受人口年龄构成因素影响的人口统计指标也同样存在功能上的效度失真问题。以总和生育率为例,全国第五次人口普查结果显示,2000年京、津、沪三大城市的总和生育率已分别降低到0.67、0.88和0.68的超低水平。这不仅大大低于目前各大城市的政策生育水平,也与各地育龄妇女的生育意愿相差悬殊。研究表明,目前中国城镇地区的总和生育率水平受“外来人口分母效应”的影响呈现显著的统计偏低现象,而人口流出地育龄妇女的总和生育率水平受“外出人口分母效应”的影响出现了明显的统计偏高现象(梁秋生,2004;郭志刚,2009)。在外来流入人口占很大比重的情况下,总和生育率指标在功能上存在明显的效度失真,既不能反映大城市育龄妇女真实的时期生育率水平,也不能反映大城市育龄妇女实际的终身生育水平。

三、在人口大流迁时期,与人口相关的区域社会经济指标的效度失真

在人口大流迁时期,不仅会造成人口流入地一系列人口统计指标的效度失真,而且会造成与人口数据相关的一系列社会经济统计指标,如人均产值、人均收入、人均住房面积等统计指标的效度失真。自2005年以来,有关这方面议题的三则新闻报道十分引人关注。

报道一^①:2005年8月,北京国际城市发展研究院、中国城市生活质量研究课题组,在

^①《外来人口,消失的人群?——我来“颠覆”城市生活质量报告》(<http://www.163.com>),2006年3月30日。

“2005 年中国城市论坛北京峰会”上发布了《中国城市生活质量报告》。这份报告首次使用“城市生活质量指数”替代 GDP 指标对中国 GDP 排名前 100 位城市的生活质量进行了排行。该《报告》发表后,在全国网民中引起强烈反响和置疑。其中置疑最为强烈的是,在城市生活质量评价过程中,是否充分考虑了城市中大量外来人口群体的生活现状。

报道二^①:2006 年 7 月 4 日,国家建设部发表的《2005 年城镇房屋概况统计公报》中公布的北京市城镇居民人均住宅建筑面积为 32.68 平方米,而北京市统计局、国家统计局北京调查总队随后公布的北京城镇居民的人均住房使用面积却仅为 19.5 平方米。两者相差 13.18 平方米。对于统计部门与建设部之间的数据差异,北京市统计局认为,其主要原因一是人均住房面积的统计口径不同。北京市统计局、国家统计局北京调查总队公布的数据是人均住房使用面积,建设部公布的是人均住房建筑面积,与统计部门测算的人均住房建筑面积 25.9 平方米仍有 6.96 平方米的差距。二是使用的人口数据不同。统计部门使用的人口为城镇建成区的常住人口,而建设部使用的人口是全市非农业户籍人口数。

报道三^②:2007 年 1 月 3 日中央电视台《新闻联播》报道,广州市人均 GDP 预计在 2006 年超过 1 万美元,从而成为中国内地首个人均 GDP 过万美元的城市。广州市统计局有关负责人随后表示,广州成为内地首个人均 GDP 过万美元的城市这个提法不确切。一是因为 GDP 增长速度的计算应剔除价格因素,而绝对值的统计未剔除价格因素,不可以简单地按增长速度去推算绝对值;二是 2005 年年末,广州全市户籍人口数达 750.53 万人,而不是 702.66 万人,加上外来人口 428.18 万人,常住人口应为 1 178.71 万人,而不是 994.3 万人;三是如果与国内其他城市的人均 GDP 相比较,应以常住人口为统一的计算口径,否则不具可比性。

三则报道的内容不同,但均涉及一个同样的人口统计口径的问题。确切地说就是在计算上述各种社会经济统计指标时,各地总人口中是否应该包括大量的外来人口群体。

无论是城市生活质量的评价指标,还是人均住房面积指标和人均 GDP 指标,都是关系到国计民生的重要统计指标,因而必然引起政府和社会公众的密切关注。所以,在计算各地区类似重大社会经济指标时,到底使用哪种统计口径的人口数据才能有效地反映真实水平,不仅是统计部门面临 的实际问题,也是人口学界必须认真对待的研究课题。

四、效度失真指标的分类及其应对方法的初步思考

在人口大流迁时期,对区域人口统计指标效度的影响是双向的,它既会对流入地人口统计指标的效度带来影响,同时也会对迁出地人口统计指标的效度带来影响,但影响的作用却截然不同。由于流迁人口对流入地来说相对集中,占流入地人口比重大,所以对流入地人口指标的影响程度较为显著。而由于人口迁出相对分散,占当地人口比重相对较小,所以对迁出地人口统计指标的影响程度也相对较小。因此,本文主要对外来流入人口较为集中的地区,特别是大城市人口统计指标效度失真的问题加以探讨。

①《北京统计局公开叫板建设部 人均住房不足 20 平米》,《上海证券报》,2006 年 7 月 10 日。

②《广州人均 GDP 过万美元续:统计局称说法不确切》,《南方日报》,2007 年 1 月 5 日。

在人口大流迁时期,受大量外来人口影响而效度失真的统计指标大致可以分为两类:一类是人口统计指标,另一类是与人口有关的社会经济统计指标。指标的类型不同,其受到的影响的性质和程度也不同,相应的解决方法和应对措施也应有所不同。在人口大流迁时期,造成流入地众多人口和社会经济指标效度失真的原因,从表面上看,似乎是现行人口统计口径变化的直接后果,但根本原因还是外来流入人口的行为后果与本地人口相比在水平或性质上存在显著的差异性。从统计学原理的角度看,外来流入人口的存在相当于在某一同类原始数据中出现了异质性的资料。所以,解决上述两类指标效度失真的问题,原则上均可以在理论分析的前提下,采用人口分类或剔除外来流入人口的方法。以京、津、沪 2000 年人口年龄构成和抚养比指标为例,把外来流入人口按年龄分布从总人口中剔除后,便可以较为真实地反映京、津、沪人口实际的年龄构成和抚养水平(见表 2)。

表 2 2000 年京、津、沪人口普查和剔除外来流入人口后的几个相关统计指标 %

	少儿系数	劳动年龄人口系数	老年系数	少儿抚养比	老年抚养比	总抚养比
北京市						
普查数据	13.59	77.99	8.42	17.42	10.80	28.22
剔除后数据	14.62	75.44	9.94	19.38	13.17	32.55
天津市						
普查数据	16.73	74.89	8.38	22.34	11.18	33.52
剔除后数据	17.27	73.86	8.87	23.73	12.18	35.91
上海市						
普查数据	12.13	76.28	11.58	15.91	15.18	31.09
剔除后数据	12.59	72.31	15.10	17.42	20.89	38.30

注:(1)2000 年京、津、沪在校大学生与研究生总数分别为 55.16 万、16.27 万、37.98 万,其中 60%按外地学生计算,并全部归入 15~64 岁人口中。(2)2000 年天津市普查数据中无外来人口年龄结构,外来人口结构按北京市外来人口年龄结构推算。

资料来源:国务院人口普查办公室、国家统计局人口和社会科技统计司:《中国 2000 年人口普查资料》,中国统计出版社,2002 年。

从表 2 数据可以看出,剔除外来流入人口后,京、津、沪的人口年龄结构指标和各项人口抚养比指标均发生了不同程度的变化,其中变化最为明显的是老年人口系数、老年抚养比和总抚养比。三市中,上海市各指标的变化幅度最大,北京市次之,天津市最小。显然,指数间变化幅度的大小与外来流入人口占常住人口的比重呈现出正相关关系。前面的分析表明,如此计算的人口年龄结构和各种抚养指标应该更能有效地反映京、津、沪 2000 年真实的人口结构和抚养水平。对于不受人口年龄结构影响的非粗人口统计指标效度失真的现象,也可以采取类似的方法。以总和生育率为例,完全可以采用剔除外来流入人口的方法计算育龄妇女的总和生育率,用以反映城镇地区实际的总和生育率水平。表 3 列出了京、津、沪 2000 年剔除外来流入人口后总和生育率的变化情况。

表 3 数据表明,大量外来流入人口的涌入是造成京、津、沪育龄妇女总和生育率统计偏

低的主要原因之一。如果剔除外来流入人口的影响因素,京、津、沪总和生育率的统计水平将会显著提高。说明在人口大流迁时期,以常住人口为基础计算的城镇地区总和生育率,既不能反映城镇地区育龄妇女真实的时期生育率水平,也不能反映城镇地区育龄妇女实际的终身生育水平。尽

管剔除外来流入人口后的指标数值更能有效说明实际问题,但在现实当中却很难获取如此精细的统计数据,这也是本文仍然使用2000年的普查数据说明问题的原因之一。在实践中,即便能够获取这样的数据,这样做也可能会引发一些理论上的争议。以剔除外来人口后计算抚养指标为例,完全否定外来流入人口对大城市人口的抚养作用就很值得探讨,因为外来人口中有相当一部分人就是直接或间接从事城市人口的照料和抚养工作的(如外来流入人口中的保姆等)。

对于解决与人口相关的社会经济统计指标的效度失真问题,在理论上似乎也完全可以采取类似处理人口统计指标的办法加以解决。但对于解决这类指标效度失真的问题,比解决人口统计指标效度失真的问题在实践上要比想象的情况复杂的多。例如,对人均收入、人均GDP这样的统计指标而言,在外来流入人口较多的地区,到底以哪种人口为基数计算这些指标更能说明问题似乎是一个很难说清的问题。因为外来流入人口对流入地的总收入和总的GDP具有重大的贡献,而且他们也把收入的一部分消费在流入地区,但他们却会同时把收入的相当一部分汇回老家抚养其留守在家的被抚养人口。所以,在计算人均收入和人均GDP时,无论是否剔除这些外来流入人口在理论上似乎都有道理,但也似乎都无道理。对于计算人均住房面积这样的统计指标时也是如此:若剔除外来流入人口,他们在流入地确实也需要住房;若不剔除外来流入人口,他们在流出地又有自己的住房,且在流入地与“本地人口”的住房需求又不一样。总之,处理这类指标的效度失真问题在理论上就存在争议。即便能够在理论上说清楚,但由于无法获取这方面详细的统计数据,也很难把外来流入人口的收入、GDP和住房面积从总体统计结果中分离出来。因此,笔者还不能在本文中给出相应的例子。

五、思考和结论

一般来说,要想正确地把握和使用某种统计指标说明和解释人口和社会经济现象的数量特征,需要具备3个基本的条件:(1)必须会正确的计算该统计指标,这是使用该指标的基本前提;(2)要了解该指标的适用性,知道使用该指标的假设条件是什么;(3)了解该指标的局限性,知道在什么情况下不能使用该指标说明问题。只有这样,才能正确无误地使用统计指标说明和解释社会经济现象的规律性。

以最常用的平均指标为例,我们必须懂得如何正确地计算各种平均指标(算术平均、几

何平均、调均、方均、中位数、众数等),并知道在什么条件下使用哪种平均数更能代表一组数据的一般水平。除此之外,我们还必须掌握各种平均数的局限性,知道在什么情况下不能使用哪种平均数。例如,在次数分布为“U”形分布、“J”形分布或极偏态次数分布的情况下,就不能使用算术平均数,而应使用其他平均数来说明问题。这就是为什么我们一般不使用算术平均年龄,而是使用年龄中位数来说明一个人口“平均年龄”的道理所在。

我们经常使用的总和生育率指标(TFR)也是如此。除了会计算该指标之外,还应知道它是一个在稳定人口的假设条件下用来表征一个人口育龄妇女一生的平均生育子女数。同时我们还应该知道,在非稳定人口条件下,总和生育率则难以正确地表征一个人口育龄妇女一生预期的平均生育子女数(马瀛通,2010)。

人口统计实践表明,不仅每一个指标均有其适用性和局限性,有时一类指标还会存在系统性的局限性,例如,几乎所有的粗人口统计指标的结果都会受人口年龄结构的影响,所以人口统计学家才设计出一系列的分年龄统计指标或标准化方法来解决这一系统性的问题。

事实上,我们目前所使用的众多人口统计指标都是在“稳定人口”的假设条件下才能真正发挥其应有的分析效度。本文的分析提醒我们,现有的许多人口统计指标不仅受限于“稳定人口”的假设条件,同样也需要在“封闭人口”的假设条件下才能真正发挥其应有的分析效度。在“开放人口”的条件下,众多人口统计指标和与人口有关的社会经济统计指标均可能会出现不同程度的效度失真现象。这是一个被我们长期忽视的系统性的人口统计学问题。

总之,在人口大流迁时期,区域人口统计指标和相关社会经济统计指标效度失真的现象是一个很值得认真研究的问题。对于从事人口学或统计学研究的学者而言,研究并解决这一问题是义不容辞的职责,而对于从事其他学科研究的学者而言,在使用人口统计指标对外来流入人口集中且规模不断扩大的城镇和地区进行相关的分析研究时,必须十分小心,否则很有可能无意中得出不科学的研究结论。到目前为止,笔者对在人口大流迁时期区域人口统计指标效度失真现象的解决办法还没有较为成熟的研究结论供读者参考,只希望本文能够起到一个抛砖引玉的作用,引起读者的重视。

参考文献:

1. 国家统计局(2006):《中国统计年鉴(2006)》,中国统计出版社。
2. 国家人口计生委(2010):《2010年实现流动人口免费服务全覆盖》(www.jiaoyitong.com),3月26日。
3. 国家统计局(2009):《中国统计年鉴(2009)》,中国统计出版社。
4. 梁秋生(2004):《外来流入人口的分母效应与大城市育龄妇女的超低总和生育率》,《人口研究》,第5期。
5. 郭志刚(2009):《近年生育率显著“回升”的由来》,《中国人口科学》,第2期。
6. 马瀛通(2010):《从稳定人口与人口再生产认识总和生育率真实涵义》,《中国人口科学》,第2期。
7. 上海市统计局(2010):《上海市2009年国民经济和社会发展统计公报》(www.city-net.cn),2月23日。
8. 北京市统计局(2010):《北京市2009年国民经济和社会发展统计公报》(www.sina.com.cn),2月2日。
9. 天津市统计局(2010):《天津市2009年国民经济和社会发展统计公报》,天津市统计信息网([www.statt-tj.gov.cn](http://statt-tj.gov.cn)),3月19日。

(责任编辑:朱犁)