

对第五次人口普查年龄结构数据的评估与调整

王广州

(中国人口与发展研究中心, 北京 100081)

摘要: 根据1982年、1990年人口普查资料和中国统计年鉴数据, 运用人口存活分析方法、人口目标分析方法对我国第五次人口普查年龄结构数据进行分析、评估和调整, 以期对深入研究“五普”数据提供参考。

关键词: 人口普查; 年龄结构; 数据质量; 数据调整

中图分类号: C923 **文献标识码:** A **文章编号:** 1000-4149(2004)05-0060-07

Evaluation and Adjustment on Age Structure Data in Fifth Census

WANG Guang-zhou

(China Population and Development Research Center, Beijing, 100081)

Abstract: Based on the population census data in 1982, 1990, and statistics year-book in 2002 the paper uses the methods of survival analysis and target analysis to analyze, evaluate and adjust the age structure data in fifth census. Hopefully, this study could provide references for further research about fifth census data.

Key words: population census; age structure; data quality; data adjustment

一、前言

年龄结构数据不仅是反映当前和过去人口状态、人口过程的最重要数据, 而且是分析未来人口过程的重要依据。分析人口年龄结构是正确估计和理解人口过程的重要方法。在人口研究过程中, 由于年龄结构数据的特殊意义与作用, 年龄结构数据质量不仅成为研究人口状态、人口过程的必要条件, 而且是研究人口问题的前提和基础。只有建立在正确人口数据基础上的研究才能有助于正确把握和理解人口问题的发生和发展过程。

随着对全国第五次人口普查(以下简称为“五普”)数据质量研究的深入, “五普”数据质量问题越来越引起研究者的高度重视。“五普”的重、漏报问题已经成为不可忽视的严重问题。尽管官方公布的“五普”漏报率高达1.81%, 即“五普”存在比较严重的漏报问题已成为无可争辩的事实。然而, 研究者们仍然对人口漏报和重报问题的严重程度产生很大的疑问^[1-2]。研究表明, “五普”数据不仅存在比较严重的漏报问题, 而且重报问题也不容忽视^[3], 正是由于重报问题的存在, 可能影响对漏报问题的判断, 这不仅关系到对人口总量的推断, 而且关系到对人口状

收稿日期: 2003-12-18; 修订日期: 2004-03-09

国家统计局, 第五次全国人口普查公报(第1号), <http://www.stats.gov.cn>, 2001-05-15 11:01:42。

作者简介: 王广州(1965-), 男, 博士, 黑龙江人, 中国人口与发展研究中心副研究员, 研究方向为人口分析方法与应用。

况、过程和特征等方面的研判与解读。研究者们反复强调的发现和析“五普”数据存在的问题不是研究的目的所在，而是发现和析“五普”数据存在的问题，纠正“五普”数据错误使之更接近实际，将误差降到最低程度。

毋庸置疑，任何调查都可能存在误差，尤其是大规模调查误差就更在所难免。从误差的种类和来源上看，误差分为系统误差和随机误差。从误差纠正的角度看，对系统误差或有规律的误差调整相对比较容易，而对随机误差的调整则比较困难。在对数据误差分析过程中首先需要对误差的种类和来源进行分析，然后，在此基础上确定误差纠正方法，最后获得比较可靠的数据。对不至影响基本结论的误差予以忽略，而对影响判断问题本质的误差则必须予以纠正。为了弥补第五次人口普查年龄结构数据的缺陷就必须对数据存在的主要问题进行全面、详细的分析与评价，然后根据已有的可靠信息对“五普”数据进行重新修正，以期达到对中国人口发展状况、发展过程全面认识和准确把握的目的，以防错误信息对人口问题判断的误导。

本文试图通过已有高质量人口普查数据对“五普”年龄结构数据存在的问题予以分析，并以此为基础试图对“五普”数据年龄结构数据进行调整，为今后研究1990年以来中国妇女的生育过程和生育水平等问题的研究奠定基础。

二、数据与方法

1. 数据

本文所采用的数据为《中国1982年人口普查资料》、《中国1990年人口普查资料》、《中国2000年人口普查资料》以及《2002年中国统计年鉴》数据。在进行数据质量分析过程中所使用的数据不包括香港、澳门和台湾。

需要说明的是为了使调整后的数据更接近实际情况，在使用上述数据对2000年人口年龄结构数据进行估计和调整时，分别对1982年现役军人年龄结构进行分解，并汇总至1982年年龄结构数据中，年龄结构分解的具体过程和方法见文献（王广州，2003）。此外，由于国家统计局《2002年中国统计年鉴》公布的各年度人口总数为年末人口数，与普查时点不同，为了使用人口总数来帮助调整和分析年龄结构，因此根据1982年以来各年年末人口数估计相应年份的年中人口数，在对年中人口数进行估计时假定人口线性增长，得到的1981~2000年全国总人口数据见表1。

表1 1981~2000年全国总人口数据表

年份		年末总人口		年中总人口	
1981	100072	99388.5	1991	115823	115078
1982	101654	100863	1992	117171	116497
1983	103008	102331	1993	118517	117844
1984	104357	103682.5	1994	119850	119183.5
1985	105851	105104	1995	121121	120485.5
1986	107507	106679	1996	122389	121755
1987	109300	108403.5	1997	123626	123007.5
1988	111026	110163	1998	124761	124193.5
1989	112704	111865	1999	125786	125273.5
1990	114333	113518.5	2000	126743	126264.5

2. 方法

人口数据质量的评估方法很多，比较常用的评估方法是通过对人口年龄结构的评估来判定人

口数据质量的高低，如年龄偏好指数等。从方法论的角度来看，常用的指数评估方法在人口数据质量评估过程中仅仅对数据的质量提供简单的评判，而无法确定问题的来源和可能产生的影响，更缺乏对数据调整具有参考价值的依据。从发现人口数据问题和计算复杂程度来看，经验指数法简便易行且只需要一次普查数据就可以进行计算，但由于多数指数基本假设条件太强且评判结果比较模糊或弹性空间较大，因此评判结果很难把握，同时很难作为比较的依据。通过计算“五普”数据质量的经验指数来看（见表2），“五普”与“三普”的惠普尔指数、迈耶尔指数、年龄准确指数和联合国年龄-性别指数非常接近。而根据经验指数的取值范围和评判标准来看，“四普”数据质量好于“五普”和“三普”，奇怪的是“五普”数据质量经验指数却与“三普”接近。显然这只是表面现象，经验指数所反映的结果与事后普查数据质量评估调查结果和相关研究不符。反观经验指数的计算方法可以看到最需要评价和最容易产生问题的年龄组经验指数也无能为力或作用有限，因此，我们不得不重新考虑上述经验指数的假设条件和评估的有效性是否适合中国人口数据质量评估。尤其是我们期望通过指数的大小变化分析人口总量误差的大小的目标无法实现。然而根据一次普查来计算经验指数的方法显然很难评估普查误差的大小。

表2 三次普查数据质量比较表

年份	惠普尔指数		迈耶尔指数		年龄准确指数		联合国年龄-性别指数	
	男性	女性	男性	女性	男性	女性	总人口	
不含现役军人	1982	101.92	102.14	3.03	3.01	6.05	6.19	33.65
	1990	101.07	100.86	2.01	2.02	4.48	4.89	29.63
	2000	101.73	101.57	2.08	2.08	6.84	7.06	34.61
含现役军人	1982	101.95	102.14	2.81	3.01	5.68	6.19	33.37
	1990	101.09	100.86	1.85	2.01	4.64	4.90	29.44
	2000	101.72	101.57	2.02	2.07	6.72	7.06	33.95

除了上述经验指数方法外，人口数据质量评估还可以采用存活分析和差分方法进行数据质量评估。存活率分析方法的优点是全面衡量人口数据年龄结构，缺点是需要两次普查数据，尤其是需要比较高质量的数据作为评判的依据，而差分方法则计算相对比较复杂。

数据评估是为对存在误差数据进行调整。数据调整则必须建立在对误差的大小与分布准确判断的基础上。

从人口数据调整的方法来看，常见的数据调整方法有数据平滑、间接估计和经验调整等。数据平滑方法是通过统计平均的方法试图消除异常值的影响，是一种折衷的办法；间接估计则建立在数据相互关系的基础上，通过已知准确信息来纠正误差；而经验调整则是根据以往的研究和经验根据主观判断人为修正误差，经验调整常用于纠正粗大误差。

从误差纠正的角度看，如何确信数据的准确程度是数据分析和调整的关键所在。只有建立在可靠数据基础上的推断才能将误差和错误降到最低。人口数据分析、误差纠正依据人口内在规律或数量关系来进行的。只有在充分研究人口内在的规律和数量特征，才能分析人口数据的可靠性，同时采用不同来源的数据进行相互补充和印证，最终实现对人口数据调整的目的。

为了评估和调整“五普”年龄结构数据，本文采用人口分析与统计分析相结合的方法，运用较少的数据达到对年龄结构调整的目的。具体方法可分为四个部分：人口存活分析、0岁人口性别比估计和0岁人口数量估计和时点调整四部分。在上述四个部分中，以人口总数为基准点，即假定各年度人口总数正确，并以此为出发点，结合人口存活分析确定出生人口数量，并根据性别比得到相应年份年中分性别0岁人口数；最后进行时点调整。即采用存活分析法和总量目标引导

的方法对普查间各年龄组人口数据进行估计和调整,该方法假定人口总数绝对误差、相对误差较小。通过上述各个步骤的推算以期达到对“五普”年龄结构数据进行比较全面衡量、认识和调整的目的。具体步骤及算法如下:

首先,采用“四普”或“三普”数据,对现存人口进行存活分析。存活分析的基本表达式为:

$${}_n P_{t_2}(x+n) = {}_n P_{t_1}(x) \cdot [{}_n L_{x+n} / {}_n L_x]$$

式中 x 的取值范围是 0 岁到 90 岁; ${}_n P_{t_1}(x)$ 是在 t_1 时刻年龄在 x 岁至 $x+n$ 岁的人口数; ${}_n P_{t_2}(x+n)$ 是在 t_2 时刻年龄在 $x+n$ 岁至 $x+2n$ 岁的人口数; ${}_n L(x)$ 确切年龄在 x 至 $x+n$ 队列存活人年数; ${}_n L_{x+n}$ 确切年龄在 $x+n$ 至 $x+2n$ 队列存活人年数。

第二, 0 岁人口性别比估计。两次普查数据出生性别比为端点,对普查间 0 岁人口性别比估计采取线性插值方法,得到各年度 0 岁人口性别比。

第三, 0 岁人口数量估计。 ${}_n P_{t_2}(0) = Tpop_{t_2} - \sum_{x=1}^{90} {}_n P_{t_2}(x)$; 式中 $Tpop_{t_2}$ 为 t_2 时刻的总人口数; ${}_n P_{t_2}(0)$ 为 t_2 时刻的 0 岁人口数。

第四,为了与“五普”数据进行对比和调整,对估计值进行时点调整,时点调整方法见文献^[4]。

最后,通过总量和分组的方法将推测数据与“五普”数据进行分析和比较。

需要指出的是,上述方法假定各年度年中人口总数估计值与实际人口总数的误差远远小于根据“五普”推算出的各年度年中人口总数的误差。由此可见,本文人口普查数据调整与评估方法的可靠性取决于总人口数和所使用普查人口年龄结构基础数据的准确程度。只有基础数据准确才能得到相对准确的估计数据。在此前提下,根据数据调整的结果与调查结果比较达到分析数据误差的大小和误差的分布的目的。因此,相关数据的准确程度直接影响对普查数据质量的判断。

3. 方法的有效性检验

数据调整的目的是使调整后的数据更接近实际情况。由于缺乏对数据检验的途径,因此只能根据已有数据对调整方法和方法的可行性予以检验。要想检验上述算法的可靠性必须同时满足两次高质量普查数据和较高质量的普查间人口总数。

由于目前可获得的高质量数据只有“三普”和“四普”数据,因此,本文采用“三普”和“四普”数据进行分析方法的有效性检验,检验的重点是新增人口的准确程度。

运用 1982 年人口普查数据和国家统计局公布的 1982~1990 年总人口数对 1990 年中国人口年龄结构进行估计,估计结果见图 1。从图 1 可以看到,对 1990 年年龄结构的估计值与“四普”的调查值非常吻合。具体估计值与普查值的误差分布见图 2。从误差的分布来看,误差主要分布在 35 岁以下各年龄组。例如,“四普”0~8 岁人口总数为 19768.86 万人,估计值为 19666.77 万人,低估 102 万人,相对误差为 5.19%。从总体上来看,0~90 岁各年龄组年别人口数累计误差为 1.79%。由于“三普”数据质量很高,因此年龄结构估计值的累积误差主要是由于“四普”年龄结构数据误差造成的。从年龄结构对比结果(参见图 1)和累计误差来看,由于人口总数相同以及“四普”年龄准确程度的原因,对“四普”年龄结构数据估计取得了很满意的估计效果。误差的主要原因在于每年总人口数量、年龄准确程度和两次普查的重漏报问题。

综上所述,本方法的主要目的在于纠正粗大误差,使调整数据更接近实际情况。上述估计结果的前提条件是“三普”数据完全正确,且各年度总人口和出生性别比数据正确,即每年的总人口数对新增人口数量的判断产生直接影响,那么其估计和调整结果比较准确。

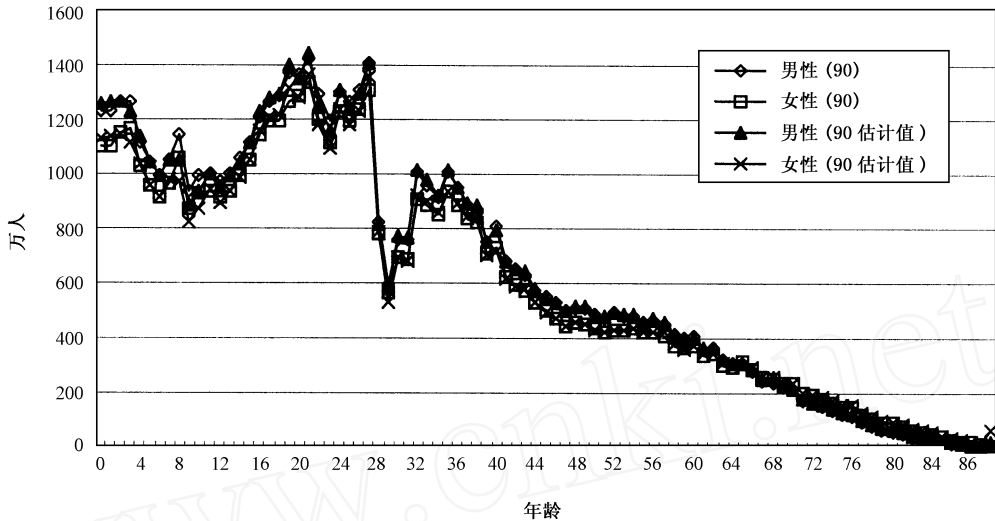


图1 1990年人口普查年龄结构分布图

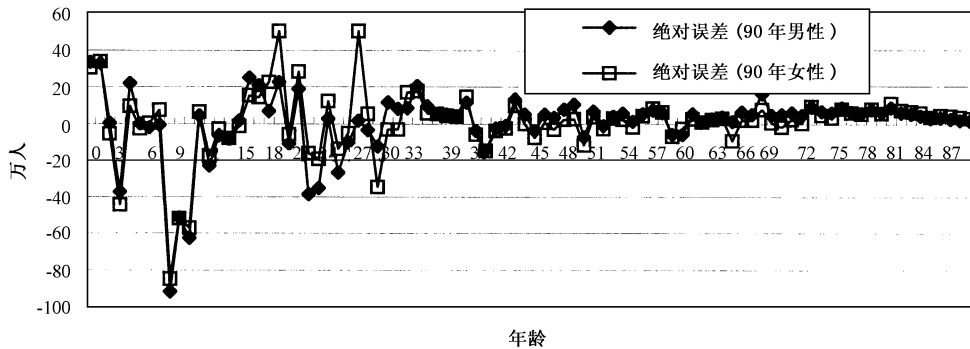


图2 1990年人口普查年龄结构与估计值误差的分布图

三、数据质量评估与调整

1. 年龄结构数据总体质量评估与结构调整

由于根据2000年人口普查计算的数据质量评估指数很难判断并确定“五普”数据存在的问题和问题的严重程度，因此本文采用“四普”和中国统计年鉴公布的总人口数和上述嫁接估计和数据调整方法，并假定国家统计局公布的各个年度总人口数据可靠，首先对2000年普查时点人口年龄结构进行估计，然后将估计值与“五普”年龄结构数据进行比较分析，判断“五普”的年龄结构可能产生的误差。

采用1990年人口普查数据和各年度总人口数对2000年普查时点人口进行估计得到2000年人口年龄结构估计结果见图3，具体年龄别人口数见表3。从估计结果看，2000年0~14岁人口占总人口的比重为25.27%；15~60岁的比重为64.70%；60岁及以上人口占总人口的比重则为10.03%。将年龄结构估计值与2000年人口普查数据比较发现，“五普”各年龄组人口与估计值的误差主要集中在0~19岁和30~40岁，尤其是0~9岁误差最大（具体误差分布见图4）。

从年龄准确程度来看，如果假定“四普”年龄结构误差为零，那么0~90岁“五普”年龄结构累计误差为7.74%，由于“五普”的重漏报远远大于“四普”，因此，在上述误差中，有理由相信“五普”错误的可能性更大。

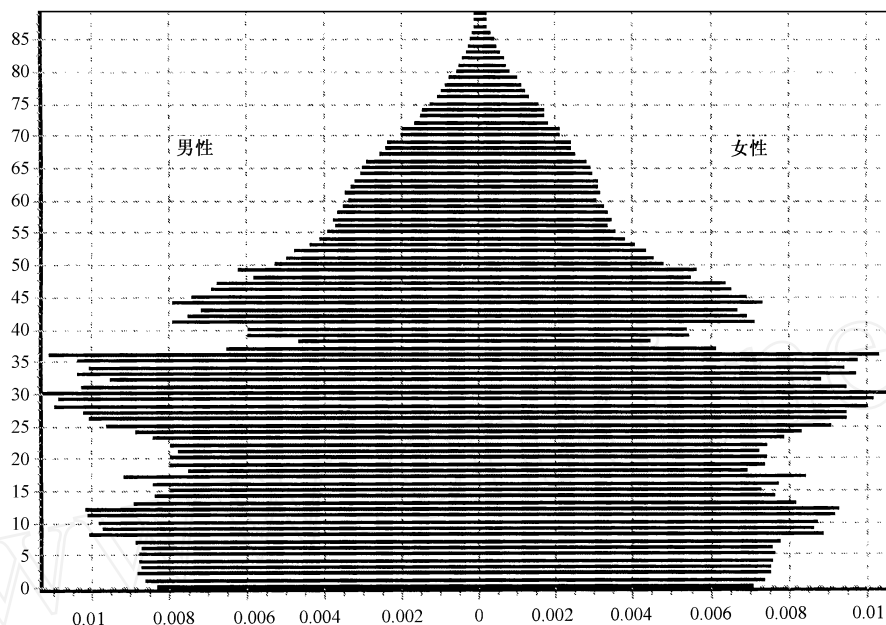


图3 2000年11月1日人口结构估计图

表3 2000年中国人口年龄结构估计值

万人

年龄	男性	女性	年龄	男性	女性
0	999.91	854.29	15~19	5123.29	4741.48
1	1028.56	881.61	20~24	4964.22	4661.46
2	1068.21	919.47	25~29	6150.04	5784.24
3	1088.5	941.14	30~34	6473.78	6079.78
4	1081.33	939.31	35~39	5289.47	5008.72
5	1083.34	945.53	40~44	4282.39	3949.31
6	1087.76	953.97	45~49	4315.04	4079.36
7	1076.88	949.04	50~54	3162.51	2925.64
8	1099.83	974	55~59	2372.39	2189.79
9	1244.05	1107.08	60~64	2088.9	1954.62
10	1205.3	1077.77	65~69	1723.63	1700.42
11	1216.03	1090.87	70~74	1183.79	1269.37
12	1255.25	1141.72	75~79	690.21	872.87
13	1257.21	1154.2	80~84	314.68	483.36
14	1105.52	1021.02	85~89	109.33	212.37
			>=90	10.52	26.2

2. 对2000年人口普查漏报问题的再认识

根据图4可以看到，“五普”低年龄组的人口漏报问题，而15~40岁主要表现为重报问题。如果在所使用基础数据中各年度人口总数正确且不考虑“四普”重漏报问题，即视“四普”重漏报人口为零，那么“五普”0~14岁人口至少漏报3300万以上，15~40岁重报数量在1100万以上。那么，“五普”的漏报率至少在2.6%以上。之所以没有按单岁年龄组进行对比分析重漏报

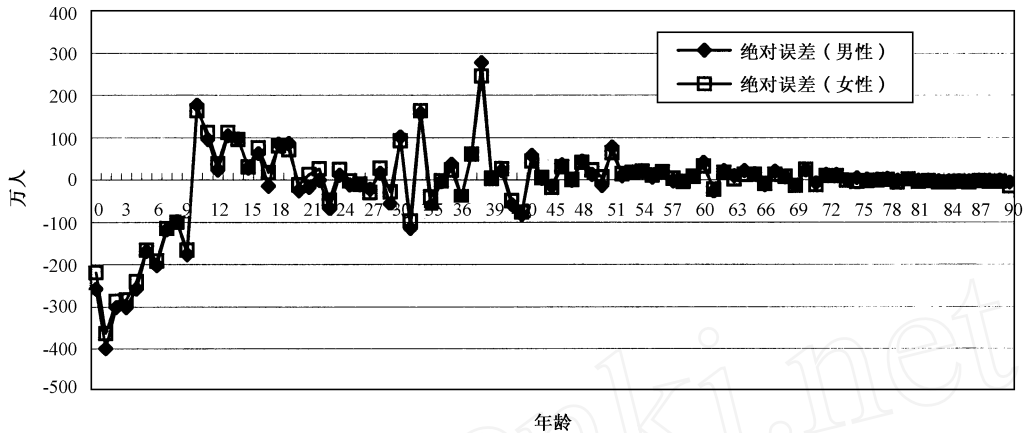


图4 2000年普查年龄别人口数与估计值误差分布图

问题，是由于年龄误报和人口出生年月年度分布等时点调整问题的存在，即为了在数据分析过程中尽可能降低由于年龄不准确带来的对总量数据的估计误差。

最后需要说明的是：本项研究试图为“五普”年龄结构数据调整方法提供思路并进行尝试，使调整结果尽可能接近实际年龄结构，在上述数据分析和调整过程中一个重要的前提是各年度总人口数相对准确，如果总人口数误差很大，那么上述研究将会出现很大偏差，因此，如何确定人口总数的正确性和可靠性将是一项关系到整个“五普”数据质量的重要问题。此外，由于“四普”人口漏报可能引起的估计误差主要表现在对2000年9岁组人口数量的高估和9岁以上年龄组低估的问题。

参考文献：

- [1] 乔晓春. 从“主要数据公报”看第五次人口普查存在的问题. 中国人口科学, 2002, (4).
- [2] 王广州. 对第五次人口普查数据重报问题的分析. 中国人口科学, 2003, (1).
- [3] 同 [2].
- [4] 同 [2].

[责任编辑 王树新]

(上接第30页)

措施。对于他们中间出现的意外怀孕，不能随意指责或埋怨，而应该给予安慰和思想疏导，使他们切实感受到政府的温暖和关心。群众相互之间的服务和管理是最有效的管理。因此要充分发挥计划生育协会在管理中的作用，通过协会会员准确掌握信息，及时提供服务，实行居民自治。

企业破产，把企业原来承担的那部分计划生育经费移交到了政府。从过去的情况看，城市区的计划生育事业经费是偏紧的。因为在计划生育经费投入要求上，对县和城市区是同一个标准，但是城市区实际负担要比农村县大。这主要是按照《条例》落实对“独生子女”家庭的优惠方面的开支。岳阳市领取了“独生子女证”的家庭共有16.3万多个，其中岳阳楼区就有5.9万多个，占36%。按照《条例》规定兑现独生子女保健费，岳阳楼区每年就需要1400多万元。原来这部分经费多数是由企业和单位承担的，再加上下岗、失业人员的节育手术费也由城市区承担，使城市区的计划生育经费开支更大，现在这笔开支中的一部分必须由政府承担。因此，城市区必须加大计划生育事业经费投入，确保对“新群体”计划生育管理的落实。

[责任编辑 王树新]