中国 90 年代以来生育水平研究[①]

中国人口与发展研究中心

生育水平的高低直接决定人口规模大小，是决定人口再生产过程和人口发展趋势的最主要的人口学因素。对于未来人口发展趋势的判断离不开对生育水平的正确估计。生育水平是否准确也关系到对劳动力人口和就业形势的估计、关系到对抚养人口比例的估计、关系到对未来人口数量和结构的估计、关系到未来人口与计划生育工作的内容和方法的转变。2000 年,中共中央和国务院颁布了《加强人口与计划生育工作稳定低生育水平的决定》,标志着我国人口发展到了一个新的阶段。但是，如果我们搞不清楚实际的生育水平，谈不上稳定低生育水平。因此，本课题的研究成果对于国家宏观社会经济政策的制定至关重要。

在得到第五次人口普查原始数据的千分之一抽样数据后,经过初步分析,我们发现第五次人口普查的原始数据基本无法使用。理由是：如果所有妇女都按照现行的国家计划生育政策规定的数量生育，总和生育率最高不过 1.55 左右,但是用第五次全国人口普查原始数据计算出的 2000 年的总和生育率只有 1.22,大大低于政策生育水平。这种结果显然是不合逻辑的，它昭示着第五次人口普查的登记人口中低龄人口漏报问题严重。如果我们利用一个错误的信息对九十年代以来的生育水平进行研究，无论方法如何科学,结果必然是错误的。于是,我们修改了原来的研究计划，将重点放到了对第五次人口普查人口年龄结构的修正上,因为只有搞清 1990 年以来出生了多少人口，才有可能估计出期间的生育水平。

本文的分为四个部分：对第五次人口普查公布的人口总数的评估、对第五次人口普查数据重报问题的分析、对第五次人口普查年龄结构数据的评估与调整、对总和生育率的估计。

一、对 90 年代以来生育水平研究的综述和评价

全国人口普查是一项庞大的社会系统工程，是在一定时点上通过对全部人口的调查,了解人口、社会、经济和其他相关信息。人口普查的结果是了解我国基本国情的宝贵资源，也是制定国民经济和社会发展，以及其他政策的重要依据。自新中国成立以来，第一次到第四次人口普查都是在计划经济体制下进行的,而 2000 年进行的第五次人口普查是在社会主义市场经济条件下完成的。同前四次人口普查相比,第五次人口普查面临的社会、经济和人口环境发生了较大变化,遇到了许多新问题,比如流动人口、人户分离、利益驱动。这些新情况,给第五次人口普查带来了前所未有的困难。经过政府的艰苦努力和人民的支持配合，圆满完成了人口普查的现场登记和复查任务。人口普查的登记不是目的，人口普查的结果才是人们关心的。因此，第五次人口普查只是完成了至关重要的一个阶段的任务，随后的数据处理和开发利用还任重而道远。目前，第五次人口普查的有关数据已经陆续公布，引起了国内外的广泛关注。其中,90 年代以来的生育水平更是人们议论的焦点之一。

妇女的生育水平是人们最关心的问题之一，因为它是直接决定人口规模大小，是决定人口再生产过程和人口发展趋势的最主要的人口学因素。生育水平是预测未来人口发展趋势的最基础信息，对于未来人口发展态势的了解离不开对生育水平的正确估计。对生育水平的判断也关系到对劳动力人口和

① 除非特殊说明外,本报告所涉及的"年代"都是指 20 世纪。

需要全本请访问 www.ertongbook.com

就业形势的估计、关系到对抚养人口比例的估计、关系到对未来人口数量和结构的估计、关系到未来人口与计划生育工作的内容和方法，等等。2000 年，中共中央和国务院颁布了《加强人口与计划生育工作稳定低生育水平的决定》，标志着我国人口发展到了一个新的阶段。但是，如果我们搞不清楚实际的生育水平，谈不上稳定低生育水平。因此，本课题的研究成果对于国家宏观社会经济政策的制定至关重要。因此，本研究具有重要的理论意义和实践意义。

测度生育水平的指标有许多，包括粗出生率、生育率、一般生育率、年龄别生育率、总和生育率、人口再生产率，等等。在实践中，到底使用哪一个或几个指标要取决于我们的研究目的和数据的可获得性。就中国的实际情况而言，使用最多的指标是粗出生率和总和生育率。人口出生率的优点在于资料容易获得，计算简单，因而几乎世界所有国家均用这个指标 其主要缺点是受人口年龄构成、性别比率、婚姻状况等因素的影响比较大。总和生育率是假定同时出生的一批妇女，按照某年的年龄别生育水平度过其一生的生育过程，每个妇女可能生育的子女数。它能综合反映各年龄组育龄妇女生育率，能确切地说明某一人群的生育水平。它不受人口性别、年龄构成和婚姻状况对生育水平的影响，故不同地区、不同年度的总和生育率可以直接比较，因而应用甚广，是测量生育水平的最好指标。

90 年代以来，各年的全国人口抽样变动调查都有人口出生率指标，我们可以从国家统计局历年出版的《中国统计年鉴》和每年的《国民经济和社会发展统计公报》中获得。从统计数据上看，90 年代以来，中国人口出生率的下降趋势非常显著，从 1990 年的 21.06‰ 下降到 2000 年的 14.03‰，10 年间大致下降了 7 个千分点。相应地，人口自然增长率也从 1990 年的 14.39‰，下降到了 2000 年的 7.58‰。尤其值得一提的是，1998 年的人口自然增长率首次在非天灾人祸的条件下降低到了 10‰以下的低水平，取得了历史性的突破，实现了老一辈国家领导人的夙愿。

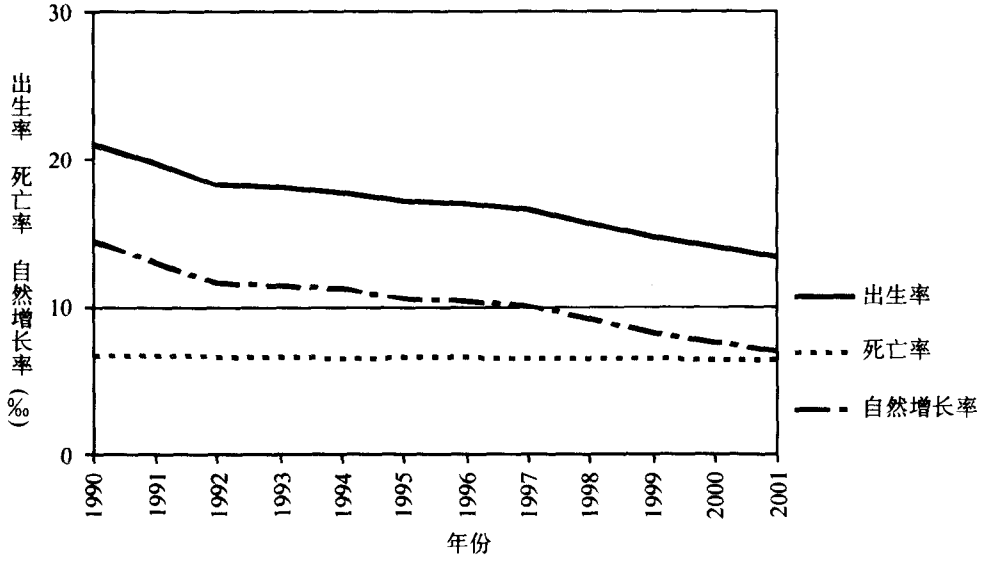

图 1　90 年代以来出生率、死亡率、自然增长率变化情况

然而，人口出生率的下降只能从一个侧面粗略地说明 90 年代中国妇女的生育水平确有下降，但是由于受人口结构的影响，人口出生率不能准确地说明 90年代以来各个年度生育水平变化的真实情况。90 年代以来，多次大型的全国抽样调查结果都显示，中国的妇女生育水平已经在冲破了 80 年代的徘徊局面，总和生育率已经降低到了更替水平以下。但是，生育水平到底比更替水平低多少？直到今天也没有一个准确而权威的说法。由于各个抽样调查的目的、调查规模、调查对象各有不同，所得到的结果也不尽相同。90 年代以来进行的这些抽样调查所显示的总和生育率有一个共同的特点：调查的妇女总和生育率大大低于实际生育水平。

例如，第四次全国人口普查结束两年后，1992 年国家计划生育委员会进行了“ 38 万人全国生育抽样调查 ”。该调查公布的全国 1991 年和 1992 年总和生育率分别为 1.65 和 1.52，大大低于总和生育率为 2.1 的替代水平。第四次人口普查所显示的总

和生育率尚在更替水平以上，一、两年内生育水平就下降了大约 1/3，显然是不能令人接受的。于是，人们自然想到出生数据漏报和瞒报是对总和生育率低估的原因。曾毅（1995 年）在人口因素分解法分析的基础上认为，1991－1992 年中国总和生育率并不象 38万人调查所报告的那样远远低于更替水平，而大致在 2.1 左右。与 1989 年的总和生育率相比，1991 年的总和生育率确有下降。下降幅度的三分之一左右是由于初婚年龄上升造成的，三分之二是由已婚妇女生育率下降造成的。

表 1　不同机构对 20世纪 90年代中国人口总和生育率的估计

调查或估计单位	1991	1992	1993	1994	1995	1996	1997	1998
中国人口信息研究中心	2.03	1.82	1.81	1.81	1.78	1.80	1.82	1.85
“38 万”抽样调查	1.65	1.52						
张二力等	2.00	1.84	1.83	1.81	1.78			
国家计生委	2.01	1.84	1.83	1.81	1.78	1.81	1.82	1.82
中国人民大学人口所	2.16	2.00	1.98	1.94	1.87	1.86	1.82	1.76
美国人口咨询局	2.2	1.9	2.0	1.9	1.8	1.8	1.8	1.8
亚太经社会	2.2	2.2	2.0	1.9	1.8	1.8	1.8	1.8

资料来源 中国人口信息研究中心，中国人民大学人口所的估计结果来自国家计划生育委员会政策法规司《未来人口发展与生育政策》研究课题报告。张二力等人的结果来自《全国和分地区人口预测》，中国人口出版社，1998 年。国家计生委的估计来自于《计划生育统计公报》(2001 年 1 号)；PRB 的结果来自《世界人口数据表，1991－99》。亚太经社会的结果来自《亚太经社会人口数据表，1991－99》。

人们不相信 38 万人调查的结果，也没有其他可以信赖的数据来源，大家期待着两次普查之间的“小普查”结果。然而，事实同样是令人失望的。1995 年由国家统计局组织的全国 1％人口抽样调查直接得到的总和生育率为 1.46，农村妇女的总和生育率只有 1.56，甚至比政策总和生育率还低 0.2。当时权威部门和有关人士估计的结果为 1.80－1.85①。1995年以后，一些全国性抽样调查实际调查的生育水平越来越低。无论从中外学者的估计看，还是从国家统计局每年公布的出生率和总人口看，人们均认为实际调查得到的生育水平并不能反映我国 90 年代真实的生育水平。由于数据的可信度和可获得性问题，国内外有关机构逐渐失去了在这方面的研究的兴趣。尽管我国有关官方机构对外称目前我国妇女总和生育率为 1.8②，但是人们无从考证其可信度到底有多大。于是，90年代以来中国妇女总和生育率成了各类统计资料的空白，也是人们猜不透的一个谜团。

2000 年，人们终于迎来了企盼已久的第五次人口普查。但是，最终的结果比以往更令人难以接受：2000 年第五次全国人口普查直接统计的总和生育率竟低至 1.22。显而易见，第五次人口普查的数据一定是在某些环节出了问题，从而导致总和生育率的失真。

面对上述调查和研究结果，如何全面、科学地分析我国 90年代以来生育水平的变化过程并得到科学的、令人信服的研究结果，成为现阶段我国生育水研究过程中的重大问题。基于上述目的本研究主要利用 90 年代以来历年国家统计局公布的有关数据，参照第三次人口普查、第四次人口普查有关数据，提出调整和修正第五次人口普查有关数据的建议，并在此基础上，研究 90 年代以来全国生育水平的变化。重点在于对 1990 年以来各年份总和生育率的估计，难点在于对人口年龄结构的调整。经过仔细研究和分析上述问题，我们的研究分以下几个步骤进行：

第一：评估第五次人口普查公布的人口总数。

第二：在肯定第五次人口普查公布的人口总数的基础上，估计重报和漏报人口数量，为理顺年龄结构作准备。

第三：在甄别重报和漏报问题的基础上，修正年龄结构。

第四：在修过的人口年龄结构的基础上，估计 90年代以来的生育水平。

① 例如，美国人口咨询局和联合国亚太经社会的估计为 1.8，张为民(1997 年)的估计为 1.85。

② 例如，中国政府在 2002年 12月曼谷《第 5 届亚太人口会议中国国家报告》中说：“1990年以来，中国的生育水平整体上呈现平稳下降态势。总和生育率从 1990 年的 2.29 下降到 2001年的 1.80”。

二、对第五次人口普查公布的人口总数的评估

中国是目前世界上人口最多的国家，快速增长的人口不利于先进生产力的发展，给中国的资源和环境带来了巨大的压力，影响了社会经济的发展，妨碍了人民生活水平的提高。正是认识到人口问题在社会经济可持续发展中的重要意义，中国政府制定“控制人口数量，提高人口素质”的人口政策，并将关系到中华民族未来的计划生育政策定为基本国策。但由于中国人口数量过多，基数过大，增长过快，直到现在，人口过多仍是我国首要问题。因此，搞清人口总数至关重要。这一艰巨的任务就是由全国人口普查来完成的。

1. 1990 年前的人口总数基本得到认可

在新中国成立之前，没有人能准确地说出中国的总人口到底是多少。人们多是沿用已经耳熟能详的数字：四万万中国人。这种认识甚至于也影响了当时的某些党和国家主要领导人。在 1949 年 9 月，毛主席就在中国人民政治协商会议第一次会议的开幕词中说，中国的人口总量已经达到了 4.75亿 。为了了解中国的基本国情，中国于 1953 年 7 月 1 日进行的第一次全国人口普查。普查数据结果表明，当年中国大陆的人口总数已经达到了 5.94 亿。毫无疑问，这一普查结果与当时人们普遍估计的人口数量有相当大的差异。第二次、第三次和第四次全国人口普查显示，1964 年、1982 年和 1990 年我国人口总数分别为 6.94 亿、10.08 亿和 11.34 亿。相对而言，第一次到第四次人口普查所得到的数据质量得到了国内外的肯定。特别是 1982 年进行的第三次人口普查所得到的数据，无论是人口总量和人口结构都受到国内外有关机构和相关学者的普遍赞誉。然而，从第四次人口普查开始，数据质量问题就开始引起人们的关注。实践也证明，人们的担心不是没有道理的。1990 年人口普查直接调查得到的人口总数比当时国家统计局根据历年抽样变动调查所掌握的人口总数少了 1500 万人。在 1990 年人口普查的数据公布后，国家统计局又不得不重新修正和调整以往曾经公布过的历年总人口数和出生率、死亡率、自然增长率的数值。

尽管人们对 1990 年第四次人口普查的人口数量以及生育水平的估计存在一些争论，但由于 80 年代进行了一系列的全国性抽样调查，包括 1982 年进行的全国千分之一生育调查、1987 年进行的全国百分之一人口抽样调查、1988 年进行的全国千分之二生育节育调查等，数据相当丰富，而且各个调查数据之间的差异基本都可以解释。最主要的是，人们对 1982 年的第三次人口普查结果信心十足，有一个可以信赖的参照数据对此后各个抽样调查所产生的误差加以合理的修正和解释。

2. 1990 年以后的人口总数遭到质疑

2000 年进行的第五次人口普查是在社会主义市场经济条件下完成的。同前四次人口普查相比，第五次人口普查面临的社会、经济和人口环境发生了较大变化，遇到了许多新问题，比如流动人口、人户分离、利益驱动。这些新情况，给第五次人口普查带来了前所未有的困难。如同 1982 年第三次人口普查后一样，1990 年以后我国也进行了多次全国性的人口抽样调查，其中包括 1992 年国家计生委进行的 38万人抽样调查、1995 年国家统计局进行的“小普查”和 1997 年国家计生委进行的生殖健康和计划生育抽样调查等。但是，这些抽样调查数据没有一个能得到大家的肯定，人们一致的观点是：漏报严重。

直到 2000 年，人们盼来了第五次全国人口普查。但是，普查开始前的摸底工作就为整个普查工作蒙上了一层阴影，各种漏报和重报的报道就不绝于耳。2001 年 3 月，第五次人口普查的公报一经见诸报端，立即引起学者们的猜疑，人们讨论的焦点集中在人口总量和人口结构上(于学军，2002 年；汪 广州，2002 年；乔晓春，2002 年；崔红艳，张为民，2002 年)。比如，公报说“普查登记结束后，全国统一抽取 602 个调查小区进行了登记质量的抽样调查。抽查结果，人口漏登率为 1.81%。祖国大陆 31 个省、自治区、直辖市的总人口中已包括据此计算的漏登人口”；而对其他误差，如重报误差率、性别误差率、年龄误差率、出生人口漏报率、死亡人口漏报率等情况没有任何说明。显然，国务院人口普查办对人口年龄和性别结构的准确性没有把握。

2002 年下半年中国统计出版社出版了《中国 2000 年人口普查资料》上、中、下三卷，人们发现许多不可思议的结果，比如总和生育率低到了 1.22，还比如 1990 年登记的 0－10 岁的人口在 2000 年第五次人口普查时突然冒出了 1000 多万，等等。显然，第五次人口普查调查到的人口结构有问题，需要

根据历年掌握的信息，用适当的技术加以修正。那么,人们也要问 第五次人口普查的人口总数是否可信？

众所周知，人口普查最大的意义在于摸清全国的人口总数，这也是任何其他全国性抽样调查所无法代替的作用。因为人口抽样调查是依据对部分人口的调查结果来推断总体各类平均指标和相对指标，而得不出总体的绝对数指标值。我国每年进行的人口变动抽样调查得到的只是调查的出生率，死亡率和自然增长率 从 1990 年以来国家分布的各年年底总人口都是依据 1990 年人口普查得到的总人口，再根据人口抽样变动调查得到的各年人口自然增长率逐年推出来的,而不是直接调查的结果。

我们注意到，2000 年第五次人口普查以后，国家统计局同样对 1990 年以来的数据进行了相应的调 整。与 1990 年第四次人口普查时调查人口数少于历年公布人口数的情况相反，这次是普查公布数多于历年公布的人口数,因此只是调整了 1998 年和 1999 年两年的人口总数，即把 1998 年总人口从 124810 万人下调到 124761 万人,减少 49 万人 把 1999 年总人口从 125909 万人下调到 125786 万人，减少 123 万人,两年合计下调 172 万人,其他年份没有变化。与 1990 年第四次人口普查相比,调整的幅度是很小的。这是否意味着国家统计局对历年人口变动抽样调查结果和公布的结果比较有把握？从而对 2000 年人口普查的总人口数量比较满意？我们的研究也正是要回答这一问题。

2001 年 3 月公布的《2000 年第五次人口普查主要数据公报》第一号显示,祖国大陆 31 个省、自治区、直辖市和现役军人的人口为 126583 万人。漏登率为 1.81%。也就是说,实际普查登记只有 124337 万人,比公布人数少 2246 万人。公布的人口总数是实际普查到的人口与漏报人口之和。问题的焦点就在于在调查到的 124337 万人基础上，加入 2246 万人是多了,还少了,还是差不多 我们不妨从以下几个信息渠道考察一下。

3.从国家统计局历年的人口变动抽样调查结果看人扣总数

我们知道,新中国成立 50 年以来,除了 1953 年、1964 年、1982 年、1990 年和 2000 年 5 个年份的人口总数是全国人口普查直接调查数据，其他所有年份的人口总数都是根据抽样变动调查数据推算的结 果。表 2 显示了 1990 年以来各年份国家统计局所掌握和公布的人口总数。从 1990 年以来各年公布的人口总数来看，各年份的人口总数与 2000 年第五次人口普查公布的人口总数在逻辑上是说得过去的,没有大起大落,人口增长率基本趋势是稳中有降。

表 2　国家统计局数据分析

年份	国家公布人口总数(万人)	省市区报表出生率(‰)	国家公布出生率(‰)	省市区报表出生(万人)	省市区公布出生数(万人)	国家公布出生数(万人)	省市区报表出生数和国家公布出生数之差(万人)
1990	114333	20.20	21.06	2308.86	2309	2407.16	98.30
1991	115823	18.32	19.68	2121.46	2121	2278.94	157.49
1992	117171	16.35	18.24	1916.22	1916	2137.73	221.51
1993	118517	15.58	18.09	1846.23	1932	2143.67	297.43
1994	119850	15.32	17.70	1835.34	1906	2120.46	285.12
1995	121121	14.42	17.12	1746.26	1833	2073.23	326.97
1996	122389	14.32	16.98	1752.77	1815	2078.35	325.58
1997	123626	13.47	16.57	1664.89	1763	2048.05	383.16
1998	124810	13.28	16.03	1657.34	1726	2000.54	343.20
1999	125909	12.70	15.23	1598.93	1662	1917.46	318.53

资料来源 中国人口信息研究中心,《人口与计划生育常用数据手册》。

但是,无论是历年公布的人口总数,还是 2000 年普查公布的人口总数都是经过人为‘ 处理 ’后的数量。如上所述,“ 五普 ”公布的人口总数加入了 2246 万人，历年公布的人口总数又是如何得到的？表 2 告诉了我们历年来各省市区抽样变动调查的结果与国家统计局经过调整后正式的公布结果的差异。从表中我们可以看到，即使国家统计局每年组织各省市区进行的人口变动抽样调查，所得到的出生人口

需要全本请访问 www.ertongbook.com

数量的结果也是偏低的,比如 1993 年和 1994 年调查得到的出生率分别为 15.58‰和 15.32‰,实际公布的结果上调为 18.09‰和 17.70‰。仅 1993 年和 1994 年两个年份就将各省市区报表的出生人口数量向上调整了 583 万。1990 年后,国家统计局每年将人口出生率加 2－3 个千分点,即平均每年加 275 万人口,10 年累积"加水"2750 万。这一数字高于第五次人口普查漏报的 2246 万人。鉴于这种"冒水"现象,国家统计局在"五普"后,向下调整了过去公布的 1998 和 1999 年的人口总数。

从这个角度看,第五次人口普查公布的的人口总数已经充分地考虑了各年各省市区上报的出生人口和人口总数的水分。国家统计局向上调整出生人口数的依据是抽查的漏报率,现在的问题是,各省市区是否在上报出生人口和人口总数时也作了向上的调整?倘若各省市区也如国家统计局一样,在汇总个地市的数据后,也作了相应的调整,那么,国家统计局所掌握的历年出生人口处和人口总数或许会高于实际的数量。这种担心不是没有道理的,表 1 告诉我们,各省市区上报的出生人口数与其公布的数量并不一致,同样作了一定幅度的上调。从表 1 我们可以看到,1990－1993 年,各省市区的报表出生人口数与公布的出生人口数完全一致,此后公布的出生人口数均大于报表出生人口数。到 1999 年,累计多出 535 万人。

总之,我们认为,第五次人口普查公布的人口总数没有超出经过"加水"后的历年人口抽样变动的调查结果,或者说,第五次人口普查的人口漏报率与国家统计局所掌握的历年各省市区上报的人口漏报率大体相当。从这个角度说,第五次人口普查公布的人口总数是可信的。

4.从国家计生委掌握的出生报表看人口总数

除了国家统计局外,掌握出生人口信息较全面的机构就是国家计生委。由于比较严格的计划生育政策使得人们的生育意愿不能得到实现,于是超生后瞒报现象比较严重是众所周知的事实。国家计生委从未否认过本部门的出生统计有水分。例如,王国强(1999 年)认为:1998 年全国计划生育率实际为 72－76% 而当年国家计生委公布的计划生育率为 93.16%。王国强、王谦和郭震威(2001 年)指出:"国家计生委近几年组织对部分农村的调查发现,各地报表出生率不同程度存在水分,高的达 30%。我们分析目前计划生育率的实际水平为 80%左右,而报表的统计超过 94%"。问题在于,国家计生委统计的出生人口漏报率到底有多高?以下几个信息或许会帮助我们回答这一问题,见表 2:

第一:1992年时,国家统计局报表和国家计划生育委员会的出生报表差距有 320 万左右的差距;1999年,这一差距为 310万。1992－1999年期间二者差距有些波动,平均每年相差大约 280 万人,二者之间的差距似乎很有规律。

第二:将国家计生委报表出生数上调 35% －40%,与国家统计局的公布出生数非常接近。

第三:2000 年,国家计生委根据报表掌握的总人口数为 12.35 亿,当年出生人口为 1292 万人[①]。根据报表出生人口数推算的总和生育率大约为 1.25,比第五次人口普查直接调查的总和生育率略高。

表 3 国家统计局和国家计生委出生人口数据比较分析

年份	SFPC 报表	SSB 公步	SFPC＋30%	SFPC＋35%	SFPC＋40%
1992	1596	2137.73	2074.80	2154.60	2234.40
1993	1568	2143.67	2038.40	2116.80	2195.20
1994	1574	2120.46	2046.20	2124.90	2203.60
1995	1521	2073.23	1977.30	2053.35	2129.40
1996	1453	2078.35	1888.90	1961.55	2034.20
1997	1388	2048.05	1804.40	1873.80	1943.20
1998	1383	2000.54	1797.90	1867.05	1936.20
1999	1287	1917.46	1673.10	1737.45	1801.80
2000	1292	1772.00	1679.60	1744.20	1808.80
2001	1240	1702.00	1612.00	1674.00	1736.00

① 《中国计划生育年鉴》2001 年卷,第 448 页。

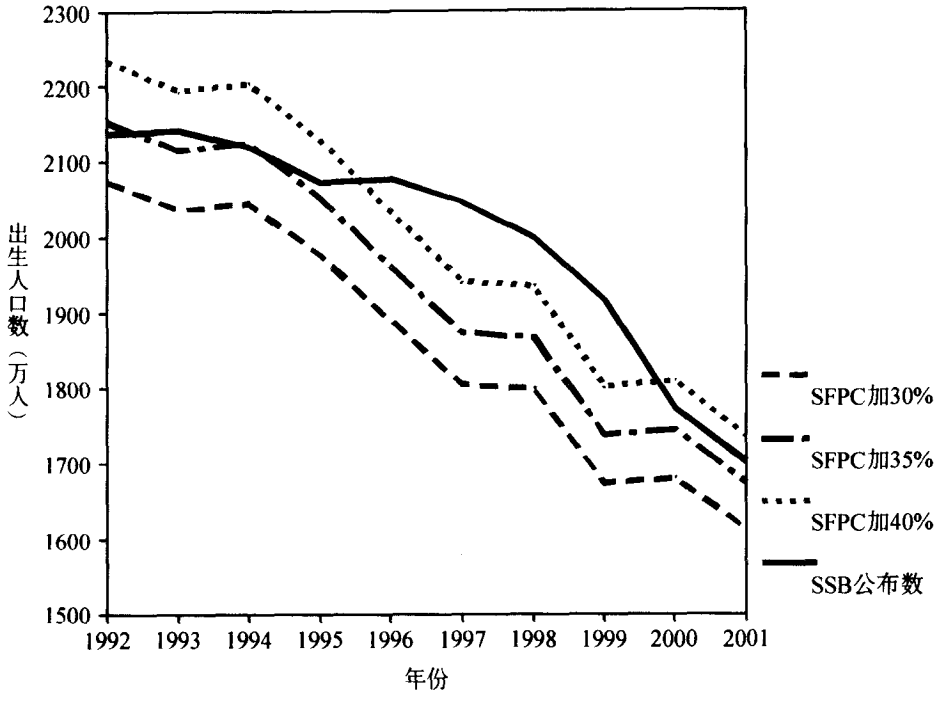

图 2 调整后国家计生委的数据与国家统计局公布的数据比较

再看国家计生委 1990年以来公布的计划生育率 。1991 年，全国国计划生育率为 80.77%，1995 年为 88.28%，而 2000 年为 94.64%。假设国家计生委报表中漏报的人口都是计划外出生的人口，在给定报表水分的前提下，我们可以推算出相应的实际计划生育率。例如，2000 年的报表计划生育率为 94.64%，如果报表水分为 30%，推算的计划生育率为 72.8% 如果报表水分为 40%，则推算的计划生育率为 67.6%。如上所述，国家计生委报表出生人口数上调 35% - 40%后，与国家统计局公布的出生人口数非常接近，换句话说，只有把国家计生委报表的计划生育率由 94.64% 下调到 70%左右之后，再以此来调整国家计生委报表出生人口数才与国家统计局公布的出生人口数大体一致。如果国家统计局公布的出生人口数据是准确的话，就可以推论出 2000 年出生人口中有 1/3是计划外超生，同时也要把国家计生委 2000 年估计的 80%左右的计划生育率再下调 10 个百分点。从历年国家计生委组织的突击抽查结果看，2000 年的报表出生人口水分高达 40%，计划生育率只有 70%的假设似乎不太可能成立。从这一点上看，国家统计局历年公布的出生人口数和人口总数应该是估计的最大值。

5. 从漏报和重报的对象看人口总数

2000 年人口普查结果中，最让人们心里没底的有三部分人口：低龄人口、流动人口和死亡人口。一般说来，0 - 10 岁的低龄人口漏报的可能性最大，而重报的可能性微乎其微。死亡人口的主要问题是漏报，而流动人口兼有重报和漏报的双重可能性。流动人口的漏报主要是受经济利益驱动的影响，用工单位故意少报外来人口或拒绝调查，主要缘于逃避劳动部门关于企业临时用工的政策，以及公安、工商、税务部门的管理。此外，由于对流动人口的各种管理收费太多，有些流动人口也不愿如实申报。武洁的研究发现，从“五普”事后质量抽查结果看，在漏报的总人口中，有近 80% 的人口是由于暂住人口漏报造成的（武洁，2002 年）。遗憾的是，文中没有进一步说明漏报的暂住人口年龄构成。流动人口的重报主要是由于离开户口所在地的流动人口既在流入地进行了登记，也在流出地进行了登记。由于信息的不对称，这部分重登人口数量无法估计。这或许是国家统计局没有公布 2000 年人口普查重报率的原因。1982 年第三次人口普查公报指出：“重报人口占 0.71‰；漏报人口占 0.56‰，重漏相抵，净差人口占 0.15‰”。1990 年第四次人口普查公报所公布的普查登记质量抽查结果是重登率为 0.1‰，漏登率为 0.7‰，重报和漏报相抵消后，净误差率为 0.6‰。但是，第五次人口普查只是简单说“普查登记结束后，全国统一抽取 602 个调查小区进行了登记质量的抽样调查。抽查结果，人口漏登率为

1.81%。”而对其他误差情况没有任何说明。是没有重报吗？不可能。与往次人口普查相比，这次人口普查面临的最大问题是流动人口大量增加，人户分离现象严重，重报现象肯定会比 10 年前增加。如果重报问题存在的话，漏报率肯定会大于公布的 1.81%。

关于漏报问题

首先让我们看以下两则消息。

“7 月 12 日，市人普办召开有渌江乡和城区五个办事处业务人员参加的紧急会议，布置出生、死亡人口摸底的复查工作，到 18 日止，全市共查出出生漏报 163 人，重报 172 人；7 月 19 日，向株洲市人普办事处报出出生、死亡人口普查数据评估报告”。（信息来源：株洲统计信息网 http://www.hntj.gov.cn）。

“公安机关在户口整顿工作中发现，全市有 13 万死亡人口未注销户口，人户分离和超生人口未入户的现象也比较严重。这次户口整顿，是为迎接全国人口普查而进行的摸底准备工作。在这次整顿中，重庆共发现 13 万死亡人口应该销户而没有销户。全市近 20% 的人口是人户分离，其中多数是外出打工等情况造成的户在人不在。整顿汇总资料还显示，全市有 6.8 万超生人口没有入户”（信息来源 新华社 2000 年 11 月 4 日）。

通过在互联网上搜索 2000 年第五次人口普查前后各地报刊上的相关内容，我们发现类似的报道很多。这些消息一方面让我们感到欣慰，它说明通过普查确实查出了一些“陈年旧帐”，另一方面我们也担心，类似的问题还有多少没有查出来？从这些报道中，我们既可以看到漏报的问题，也可以看到重报的问题。无论是漏报还是重报，都会影响人口总数。死亡人口如果没有注销户口，而且在人口抽样变动调查或人口普查时继续登记在册，也会影响人口总数。比如 1999 年底重庆市全市人口为 3072 万人，清查出来的未注销户口的死亡人口为 13 万，占总人口的 4.23‰。清查前，这部分死亡人口一直“冒充”总人口的一个组成部分，无疑会给历年公布的人口总数打一定的折扣。

至于低龄人口的漏报问题已经是存在多年的老问题。人们普遍认为，第五次人口普查登记的低龄人口同样存在严重的漏报问题，到底漏报多少，还需要进一步的研究。于学军（2002 年）根据第四次人口普查的年龄别死亡率推算出来的 2000 年第五次人口普查时 0－14 岁未成年人口数量应该为 3.22 亿，占总人口的 25.4%，而第五次人口普查官方公布的结果是：0－14 岁的人口为 2.90 亿人，占总人口的 22.89%，估计数量与官方公布的数量之间相差 3200 万人。崔红艳张为民（2002 年）依据 6－10 岁入学人数推算漏报人口后发现：仅 6－10 岁年龄段漏登人口就达 1208 万，占普查实际登记人数的 0.98%，再次说明，这次普查直接登记的总人口确实有漏报。如果再考虑漏报不可能仅仅发生在 6－10 岁儿童中，更低年龄 0－4 岁漏报的程度可能比 6－10 岁大，那么总人口的漏报超过 2000 万是完全可能的”。

关于重报问题

以上我们分析的是漏报的情况，重报的情况又如何呢？首先让我们来观察一个矛盾的现象：按照第五次人口普查结果，10－20 岁的人口为 24582 万人，平均每年出生 2235 万人。作为同一批人，他们在 1990 年第四次人口普查时相应地为 0－10 岁。按照四普结果，这批人口为 23500 万人，平均每年出生 2136 万人。这意味着，在没有考虑死亡率和两次人口普查时点差异的条件下，同样是一批人，10 年间多出了 1082 万人。出现这种矛盾的现象说明，要么第四次人口普查有问题，要么第五次人口普查有问题。

崔红艳和张为民（2002 年）认为：“这种不合理现象，有两种解释，一是 2000 年人口普查重登了 10－19 岁人口，根据 2000 年人口普查现场登记情况以及以往调查经验，这种重登的理由不充分，因此可能性不大；二是 1990 年相应年龄，即 0－9 岁年龄人口有不同程度的漏报，这种情况可能性较大”。但我们认为，这种现象可能是 2000 年重登和 1990 年漏登两种误差双重作用的结果，只是后者的作用更大一些。理由是，在 10－19 岁的人口中，有为数不少的流动人口。第五次人口普查的结果显示，2000 年我国平均外出时间超过半年以上者达 12107 万人。其中，0－9 岁的流动人口大约为 1200 万人，10－19 岁的流动人口大约为 2000 万人。

既然第五次人口普查不能告诉我们到底有多少重报人口，我们只能以相对比较可信的第三次、第四次人口普查的人口结构为基础，假设一个合理的死亡率，构造一个生命表，通过比较历史人口的年龄别存活率和第五次人口普查登记的相应年龄别人口存活率来判别重报情况。从理论上说，如果后者大于 1，则说明有重报现象。崔红艳和张为民（2002 年）所作的类似估计显示，在 10－44 岁的 35 个年龄组中，有 21 年龄组的存活率大于 1。其中 24－44 岁的 21 个年龄组中，有 12 个年龄组的存活率大于 1。

表 4　同一批人在‘四普’和‘五普’的登记数量对比

2000 五普时的年龄	1990 年四普时的年龄	出生年份	五普时申报的出生数量	四普时申报的出生数量	五普和四普申报的出生数量差额
10	0	1990	2621.00	2322.09	298.92
11	1	1989	2513.77	2333.19	180.58
12	2	1988	2457.62	2418.06	39.56
13	3	1987	2528.26	2429.44	98.83
14	4	1986	2319.01	2141.07	177.93
15	5	1985	2042.93	1998.23	44.70
16	6	1984	2031.34	1905.61	125.73
17	7	1983	2006.50	2019.65	-13.14
18	8	1982	2310.04	2201.53	108.51
19	9	1981	1912.29	1808.65	103.64
20	10	1980	1839.38	1922.41	-83.03
合计			24582	23500	1082

如果说 10－19 岁组的人口在 1990年人口普查时有不同程度的漏报，那么 24－44 岁各年龄组中有一半以上存活率大于 1 又作何解释呢？答案只有一个：主要是重报的结果。

6. 主要结论

第一，1990 年后，国家统计局将每年的人口抽样变动调查得到的人口出生率加 2－3 个千分点，即平均每年加出生人口 275 万，10 年累积‘加水’大约 2750万。第五次人口普查公布的人口总数没有超出经过‘加水’后的历年人口抽样变动的调查结果，或者说，第五次人口普查的人口漏报率与国家统计局所掌握的历年各省市区上报的人口漏报率大体相当。从这个角度说，第五次人口普查公布的人口总数是可信的。

第二，只有把国家计生委报表出生人口数上调 35％－40％，或者把国家计生委报表的计划生育率由 94.64％下调到 70％左右之后，国家计生委掌握的出生人口数才与国家统计局公布的出生人口数大体一致。从历年国家计生委组织的突击抽查结果看，2000 年的报表出生人口水分高达 40％，计划生育率只有 70％的假设似乎不太可能成立。从这一点上看，国家统计局历年公布的出生人口数和人口总数应该是估计的最大值。

第三，第五次人口普查调查的人口既有漏报，也有重报。公布的漏报率为 1.81％ 或许是重报和漏报相抵消后的净误差率。

总之，我们认为，第五次人口普查公布的总人口数为 126583万是可信的，但是人口年龄结构的误差非常严重，需要重新估计漏报率和重报率后加以相应地调整。

三、对第五次人口普查数据重报问题的分析

第五次人口普查存在比较严重的漏报问题成为一个重要的、不可忽视的问题。尽管第五次人口普查官方公布的漏报率高达 1.81％，然而，由于缺乏比较详尽的漏报人口年龄结构等数据信息，因此对漏报人口的分布和判断成为比较困难的问题。目前比较一致的看法是漏报人口大体主要落在 20 岁以下，特别是 0－9 岁儿童漏报可能是主要因素（于学军，2002；崔红艳等，2002）。由于人口普查存在比较严重的漏报问题，那就意味着对相关人口状态和过程的推断将会存在很大的偏差。人口数据分析、调整和间接估计等任务将异常繁重，尤其是对人口数量、质量、结构、水平、过程和发展趋势等一系列问题的研究和评判将直接影响到人口与计划生育政策的制定和实施，影响未来中国人口发展过程和发展趋势。同时，由于人口过程具有长周期、不可逆和滞后性，因此任何决策失误都会对未来中国社会发展导致不堪设想的严重后果。

除了漏报问题以外，‘五普’是否也存在比较严重的重报问题 遗憾的是‘五普’并没有给出可供参考的数据（乔晓春，2002），这就给正确评价‘五普’数据带来了一定的困难。可是，为了弥补人口普查数据的缺陷就必须对人口普查数据的问题所在进行全面、详细的分析和评价，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、发展过程的全面认识和准确把握,以防错误信息对人口问题判断的误导。在此我们试图对第五次人口普查存在的重报问题予以论述，以便为今后“五普”数据年龄结构调整、生育水平的判定以及死亡等问题的研究提供参考。

由于我国由计划经济向市场经济体制的转换，社会、经济、文化构成向多元化方向发展,受户籍制度改革和社会结构转型的影响，人口信息的构成和收集遇到前所未有的困难，这不仅是数据量增加带来的困难，更是由人口结构的复杂性对人口信息产生巨大影响，因此高质量数据获得难度加大。纵观我国大规模人口调查的历史，我国已进行了五次人口普查和多次大规模的人口抽样调查，人口数据信息日益丰富,尤其是 1982年人口普查数据质量之高是前所未有的。这对分析和判断以前和以后的人口普查与人口调查起到重要的作用。 1990 年人口普查也是一次较高质量的人口普查，因此上述数据对分析、研究第五次人口普查数据的质量,如数据漏报和重报问题具有重要意义。本研究所采用的数据为《中国 1982 年人口普查资料（电子计算机汇总）》，《中国 1990 年人口普查资料》以及《中国 2000 年人口普查资料》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。

为了对第五次人口普查存在的重报问题予以分析,本项研究从年龄结构分析出发,采用的方法是根据已有的高质量人口数据推算 2000 年人口状况,将推算结果与 2000 年实际普查数据进行比较分析,以期达到对“五普”数据的数据质量进行比较全面衡量和认识的目的。在分析方法上采用存活分析法、对普查时点进行调整和数据汇总比较三个步骤。

首先，存活分析方法的应用可以达到分析 2000 年人口普查中相应人口数据的准确程度的目的。存活分析的基本表达式为：

$$ {}_nP_{t_2}(x+n) = {}_nP_{t_1}(x) \cdot [{}_nL(x+n)/{}_nL(x)] $$

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

第二，由于第五次人口普查与第三次人口普查和第四次人口普查的普查时点不同，为了使存活分析更接近第五次人口普查时的人口状态，相应地将存活分析的时点调整到与第五次人口普查相同，即 11 月 1 日。各年龄组数据时点人口计算方法是：

$$ {}_nP'_{t_{11}}(x) = {}_nP_{t_1}(x) + ({}_nP_{t_2}(x+n) - {}_nP_{t_1}(x)) \cdot 0.25 $$

式中 ${}_nP'_{t_{11}}(x)$ 是在 2000 年 11 月 1 日时年龄在 x 岁至 $x+n$ 岁的人口数；${}_nP_{t_1}(x)$ 是在 2000 年 7 月 1 日年龄在 x 岁至 $x+n$ 岁的人口数 ${}_nP_{t_2}(x+n)$ 是在 2001 年 7 月 1 日时年龄在 $x+n$ 岁至 $x+2n$ 岁的人口数。

第三,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。

最后需要说明的是由于第三次人口普查现役军人年龄结构是按 5 岁分组,缺少按 1 岁分组的数据资料，但为了尽可能准确地反映普查数据间的数量关系,对第三次人口普查现役军人采取按“四普”公布现役军人年龄结构进行分解的方式得到 1982 年 1 岁组现役军人的年龄结构。由于各年龄组现役军人的相对数量较少，且现役军人年龄结构变化不会很大,因此,现役军人年龄结构按 1 岁分解后的年龄别人口数不会对总人口的年龄结构带来很大的分解误差。

对人口普查数据质量的检验事关重大，为了表明推算方法的有效性和研究问题的需要，使用高质量数据进行检验就显得尤为重要，这不仅关系到对推算准确性的评价，而且是涉及到对事后质量检验的可靠性评价。由于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。在用“三普”数据对“四普”8－90 岁人口进行存活分析时,考虑到“四普”死亡漏报问题,因此将男、女平均预期寿命分别设定 68.3 岁和 71.3 岁,对 1982 与 1990 年间各年度的男性和女性预期寿命采取线性差值的方法得到。根据上述假定的推测结果见表 5。从表 5 可以看到,根据“三普”数据推测 1990 年 8－90 岁男性人口为 49355.6万、女性人口为 46641.3万,男女合计为 95996.9 万 相应地，1990 年实际普查数男性为 49294.9、女性为 46481.0,男女合计为 95775.9 万。推测值与实际调查的差值为 221 万人,即推测数据比实际普查数多了 220 多万人。根据 1990 年质量抽查漏报率为 0.7%,重报率为 0.1%,0－100 岁及以上的人数净误差为 0.6% ,即推算可能漏报 791 万人,净误差为 678 万。因此从推算结果可以断定 1990人口普查漏报的发生，并与事后抽样结果的结论是一致的。此外,根据推测可以看到 18－60 岁数据质量相对较高,净误差为 196.0 万人,而 60 岁以

上则相对较差。由于“三普”和“四普”调查时点相同,不涉及时点调整问题,因此从推算符合精度上来看相对较高。总之,数据检验表明上述分析方法可以比较确切地反映问题所在。

表5 1990年人口普查数据与1982推测数据比较表 单位:万人

1990年人口数		1990年普查		1982年普查推算	
		男	女	男	女
8-90岁		49294.9	46481.0	49355.6	46641.3
合计		95775.9		95996.9	
其中	18-90岁	38620.7	36453.0	38873.4	36799.6
	合计	75073.7		75673.0	
	18-60岁	34415.5	31762.0	34470.0	31903.5
	合计	66177.5		66373.5	
	61-90	4205.2	4691.0	4403.4	4896.1
	合计	8896.2		9299.5	

由于我们可以获得的准确和比较准确的近期人口普查数据为1982年人口普查和1990年人口普查数据,虽然从时间的有效性的角度看,1990年人口普查数据好于1982年人口普查,但由于1990年普查也存在一定的人口漏报问题,漏报率达0.7%,相比之下1982年人口普查漏报率仅0.56‰,因此从数据的有效性和完整性上来看,1982年人口普查是一个难得的准确数据标准。此外,由于死亡水平变化相对比较缓慢且方向确定,而且对总人口的变化影响相对比较确切,因此通过恰当的假定对现存人口的预测误差不会很大。正是出于上述考虑,为了比较全面验证和推断2000年人口普查数据存在的主要问题,下面研究采取1982年人口普查为主、1990年人口普查为辅,结合部分2000年人口普查数据的分析策略来分析“五普”的重报问题。

根据2000年人口普查数据计算男性和女性人口的预期寿命分别为70.61和74.45岁,考虑到死亡漏报和男性和女性预期寿命的差距问题,因此将2000年男性和女性预期寿命估计为70和73岁,对1982或1990年与2000年间各年度的男性和女性预期寿命同样采取线性差值的方法得到。基于上述数据、算法和基本思路,得到2000年11月1日中国人口18-90岁推测值,具体数值见表6。

表6 2000年人口普查数据与人口推测值比较表 单位:万人

		2000年普查		1982年普查推算		1990年普查推算	
		男	女	男	女	男	女
10-90岁		55649.0	52880.5	-	-	54384.0	51472.5
合计		108529.5		-		105856.5	
其中	18-90岁	45978.0	43928.7	45412.2	43256.1	45281.6	43164.2
	合计	89906.7		88668.3		88445.8	
	18-60岁	40136.5	37755.1	39466.0	36942.3	39546.1	37019.8
	合计	77891.6		76408.3		76565.9	
	61-90岁	5841.5	6173.6	5946.2	6313.9	5735.5	6144.4
	合计	12015.1		12260.1		11879.9	

从表6可以看到,与1982年的推算的结果相比,2000年人口普查得到18-90岁的人口数量多了1238.4万,其中,18-60岁多了1483.3万,而61-90岁则少了244.9万。如果不考虑时点转换问题,直接用1982年数据推算到2000年7月1日,18-90岁推算结果是男性45525.9万人,女性43353.7万人,合计88879.6万人,相应地,与2000年人口普查的误差为1027万人。然而,事后人口普查数据质量抽样结果认为,2000年人口普查漏报率为1.8%,即漏报2246万人,比较“四普”漏报情况,这里推算的结果与实际调查结果存在数据矛盾现象。如果根据1990年人口普查数据推算,与推算结果比较,10-90岁2000年普查多报2673万人,其中18-90岁多报1461万人。如上所述,由于根据1990年人口普查质量抽查的漏报率0.7%推算,1990年可能漏报791万人,那么,如果假定1990年所有漏报人口都发生在0-80岁之间(80岁以上人口比重相对较小,忽略不计),可以推断10-90岁2000年普查至少多报1800万人。鉴于上述推测可以确信2000年人口普查存在比较严重的重报问题。如果不考虑

需要全本请访问 www.ertongbook.com

10－90 岁人口的重漏与漏报的相抵问题，年龄结构数据的差错率可能会相当大。

经过使用 1982 年人口数据和 1990 年人口数据的推算值与 2000 年人口普查的可以看到人口重报问题的存在。由于 1982 年人口普查数据资料质量相对较高，可以以此作为数据推测与判断的依据，通过比较推测值和实际普查值看到“五普”误差的分布情况(见图 3)。从图 3 可以看到，重报问题主要发生在 45 岁以下，尤其是 30－40 岁和 20 岁以下。如 30－40 岁重报人口占全部重报人口一半以上，具体情况见表 7。从表 7 可以看到误差最大的年龄组依次为 38、30、31、32，尤其是 37、38、39 和 40 岁连续出现负值，即推算低估现象，而 1982 年对 1990 年进行推测误差却没有发生连续为负值或正值的情况，可以断定 2000 年普查重报现象的发生，而绝非运算方法带来的误差。

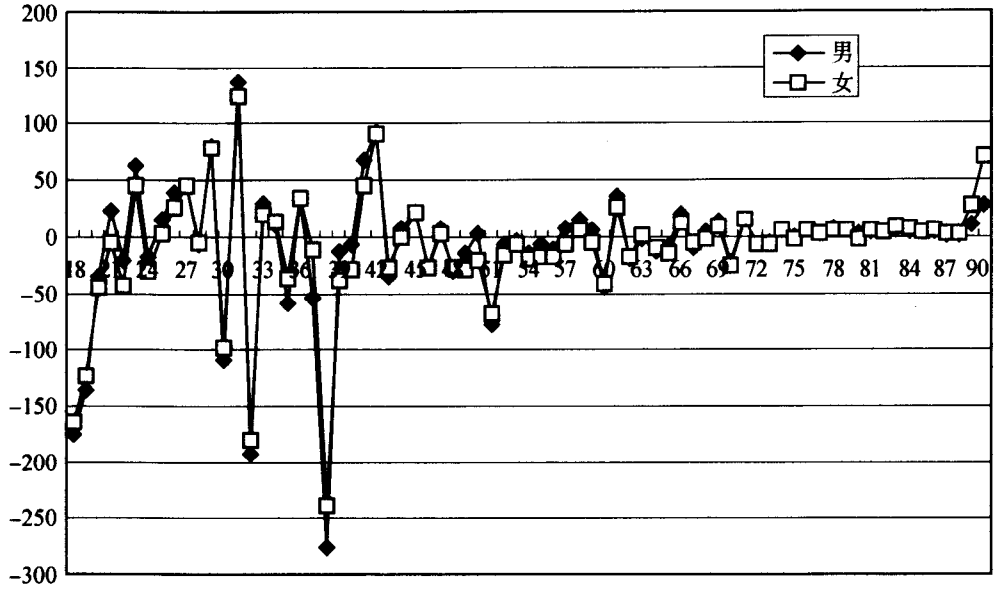

图 3　2000 年人口普查数据与 1982 年普查数据推算值误差分布图①　　单位 万人

表 7　2000 年人口推测值与人口普查数据误差比较表②　　单位 万人

年龄	2000 年普查人口数		2000 年人口推测值		估计误差	
	男	女	男	女	男	女
30	1444.0	1362.0	1334.7	1263.7	－109.3	－98.3
31	1286.0	1220.5	1422.6	1345.5	136.6	125
32	1429.3	1345.9	1237.0	1166.2	－192.3	－179.7
33	1116.3	1060.6	1145.2	1081.2	28.9	20.6
34	1278.3	1208.1	1288.8	1222.8	10.5	14.7
35	1282.9	1200.7	1224.4	1164.5	－58.5	－36.2
36	1241.1	1177.4	1275.6	1212.4	34.5	35
37	1439.8	1350.3	1385.1	1338.7	－54.7	－11.6
38	1081.5	1012.6	804.6	773.4	－276.9	－239.2
39	581.8	561.0	568.5	523.6	－13.3	－37.4
40	762.0	707.7	754.6	678.7	－7.4	－29
合计	12943.0	12206.8	12441.1	11770.7	－501.9	－436.1
	25149.8		24211.8		－938	

① 根据 1982 年人口普查数据推算。误差＝1982 年预测值－2000 年普查值。

② 同上。

如果“五普”10－90岁人口重报高达1800万以上的话,那么可以确信总人口的重报会更大一些,因此，我们不得不对总人口数量和人口漏报问题重新审视。假定普查事后抽样漏报率准确，且漏报全部发生在10－90岁年龄组,那么，10－90岁年龄组人口重报将高达4000万,总人口将在12.26亿左右。如果假定漏报全部发生在0－9岁年龄组,人口重报也将高达1800多万,那么,总人口可能将在12.48亿左右。由此可见,如果普查事后抽样漏报率准确，那么2000年人口普查中人口重报数量可能在1800万至4000万之间，总人口数量则应该在12.26至12.48亿之间。如果普查事后抽样漏报率存在偏差,那么“五普”的总人口数量将无从判断。

最后需要说明的是，虽然上述分析可以断定人口重报问题的发生和问题的严重性，但由于对死亡水平假定和年龄结构时点调整可能影响推断的精度。无论精度如何，如果上述重报问题存在且比较严重,那么不仅影响到对2000年总人口的推断,而且几乎需要对所有数据进行重新调整和修正，尤其是对0－9岁年龄组数据的修正直接关系到对中国人口生育水平的判断和未来人口政策的制定。

四、对第五次人口普查年龄结构数据的评估与调整

年龄结构数据质量不仅是反映人口状态和人口过程的最重要数据，而且也是影响未来人口过程的重要基础。分析人口年龄结构数据是正确估计和理解人口过程的重要手段。正是由于年龄结构数据的特殊意义与作用，年龄结构数据质量的高低不仅是研究人口状态和人口过程的必要条件，而且是研究人口问题的前提和基础。只有建立在正确数据信息基础上的研究才能正确把握和理解人口问题的发生和发展过程。

随着对第五次人口普查(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和影响数据质量的重要问题。尽管第五次人口普查官方公布的漏报率高达1.81%,即第五次人口普查存在比较严重的漏报问题已成为无可争辩的事实。然而，研究者仍然对人口漏报和重报问题的严重程度产生很大的疑问(乔晓春，2002;王广州，2003)。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（王广州，2003),正是由于重报问题的存在,可能影响对漏报问题的判断，这不仅关系到对我国总人口数量的推断,而且关系到对中国人口状况、过程和特征等方方面面问题的研判与解读。

正如研究者们反复强调的是，发现和分析“五普”数据存在的问题不是研究的目的所在。发现和分析“五普”数据存在的问题是正确使用和纠正“五普”数据错误过程中必须进行且无法逾越的重要阶段。诚然,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。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，以此为基础确定误差数据调整的方法。对于不至于影响基本结论的误差予以忽略，而对于可能影响对问题本质判断的误差则必须根据相关准确信息予以调整和纠正。为了弥补第五次人口普查年龄结构数据的缺陷就必须对人口普查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、发展过程的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。

在此，我们试图通过已有高质量人口普查数据对第五次人口普查年龄结构数据存在的问题予以论述,并依次为基础试图对“五普”数据年龄结构数据进行调整，为今后研究九十年代以来中国育龄妇女的生育过程和生育水平等问题的奠定研究基础，而本文的分析重点将放到对“五普”数据的评估与调整上。

我们所采用的数据为《中国1982年人口普查资料(电子计算机汇总)》,《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，使各种可能的误差降到最低。因此对误差的产生渠道予以分析，以便确定估计方法。因此在使用上述数据对2000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，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

来各年年末人口数对估计相应年份年中人口数，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 全国总人口数据具体数据见表 8。

表 8　1981－2000 全国总人口数据表单位：万人

年份	年末总人口	年中总人口	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

人口数据质量评估与调整是人口研究的重中之重。人口数据质量的评估方法很多。比较常用的人口数据质量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。在对人口数据质量评估过程中，本项研究首先采用常用人口普查数据质量的评估方法，以便确定人口数据存在的主要问题。之所以采用常用人口数据质量评估方法，目的在于与以往的普查数据质量进行比较分析。然而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，更缺乏调整数据的依据。从人口数据的分析和研究的角度来看，对普查数据的研究仅停留在发现问题上是远远不够的，更重要的目的在于问题的解决。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，但评判结果弹性空间较大，比较模糊。通过计算“五普”数据质量的经验指数来看(见表 9)，“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－性别指数的非常接近。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”，“五普”数据质量与“三普”接近。这只是表面现象，显然与事后普查数据质量评估调查结果和相关研究不符。因此，我们不得不重新考虑上述经验指数的有效性。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。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 9　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等进行数据质量评估。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。

如前所述，数据评估并不是研究的根本目的所在，对存在误差数据的调整才是研究的最终目标。数据调整是对存在偏差人口数据进行分析和研究的重要环节。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、误差纠正是依据人口内在规律或数量关系来进行的。只有根据人口内在的规律和数量特征研究和分析数据的可靠性，同时采用不同来源的数据相互补充和印证，才能对人口年龄结构数据的调整，因此修正原有数据的误差是人口数据调整的重要任务所在。

为了评估和调整“五普”年龄结构数据，我们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。具体方法分为三个部分：人口存活分析0岁人口性别比估计和0岁人口数量估计三部分。在上述三个步骤中，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。也就是采用存活分析法和总量目标引导的方法对普查间地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。通过上述步骤以期达到对“五普”年龄结构数据进行比较全面衡量、认识和调整的目的，其具体算法如下：

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：

$$ {}_nP_{t_2}(x+n) = {}_nP_{t_1}(x)\cdot[{}_nL(x+n)/{}_nL(x)] $$

式中 x 的取值范围是0岁到90岁；${}_nP_{t_1}(x)$ 是在 t_1 时刻年龄在 x 岁至 $x+n$ 岁的人口数；${}_nP_{t_2}(x+n)$ 是在 t_2 时刻年龄在 $x+n$ 岁至 $x+2n$ 岁的人口数；${}_nL(x)$ 确切年龄在 x 至 $x+n$ 队列存活人年数；${}_nL(x+n)$ 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。${}_nP_{t_2}(0) = Tpop_{t_2} - \sum_{x=1}^{90} {}_nP_{t_2}(x)$；式中 $Tpop_{t_2w}$ 为 t_2 时刻的总人口数；${}_nP_{t_2}(0)$ 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(王广州，2003)。

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总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，上述人口普查数据调整与评估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。由此可见，相关数据的准确程度直接影响对普查数据质量的判断。

数据调整的目的是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年总人口数数据对1990年中国人口年龄结构进行估计，估计结果见图4。从图4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5 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岁年别人口数累计误差为1.79% 根据1990年质量抽查漏报率为0.7%，重报率为0.1% 年龄误差为3.07%。1982年人口普查漏报率仅0.56‰，重报率为0.71‰，年龄误差为6.15‰。如从年龄准确程度来考虑，由于两次普查误差的原因，0－100岁年龄别人口数累计误差不可能低于2.445%－3.685%年龄误差范围。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或许是由于“四普”年龄结构数据误差造成的。因此，从年龄结构对比结果和累计误差来看，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，其误差是完全可以接受的。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。每年的总人口对新增人口数量的判断产生直接影响。

综上所述，本方法在于纠正粗大误差，使调整数

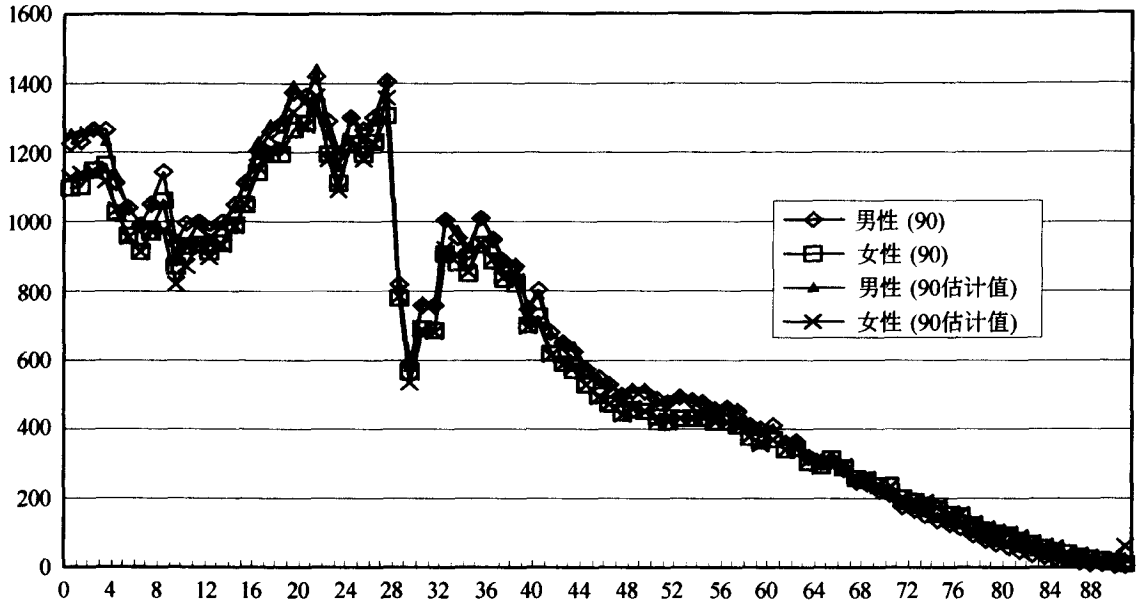

图 4　1990年人口普查年龄结构分布图

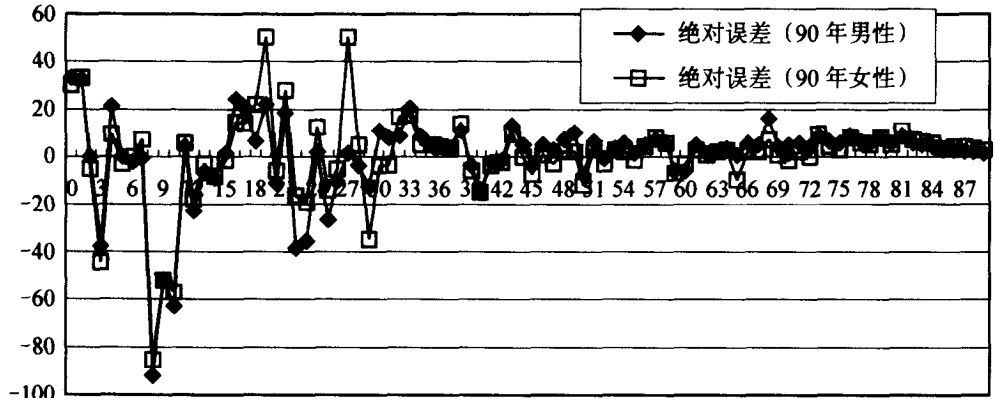

图 5　1990年人口普查年龄结构与估计值误差的分布图

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,那么其估计和调整结果比较准确。

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”数据进行比较分析。

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,2000年人口年龄结构估计结果见图6,具体年龄别人口数见表10。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7)。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%。如果在上述累计误差中,一部分原因完全是由“四普”误差造成的,即如果消除“四普”年龄结构数据的误差可能产生估计误差问题，那么,“五普”年龄结构累计误差也应该在4.5%以上,即“五普”的年龄结构数据总误差在4.5%以上。如果上述误差中“四普”年龄结构误差只有一半的误差“传递”到“五普”,那么“五普”年龄结构累计误差也应该在6.2%以上。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。

根据图7可以看到“五普”从总体上看,低年龄组主要表现为漏报问题,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

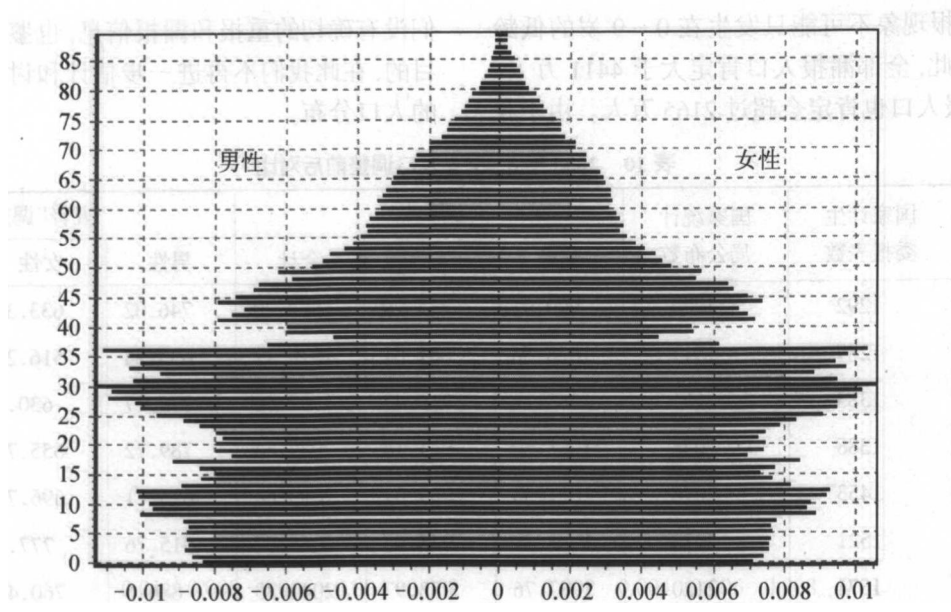

图 6　2000 年 11 月 1 日人口结构估计图

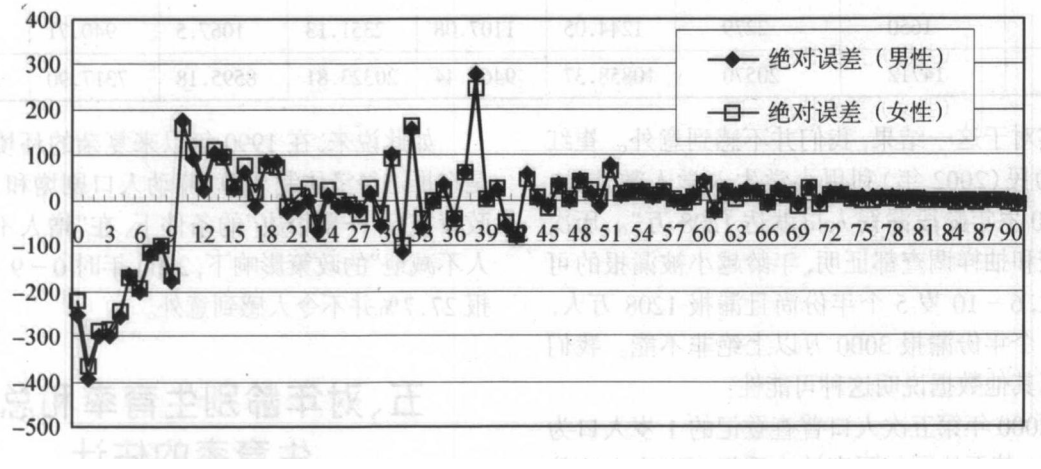

图 7　2000 年普查年龄别人口数与估计值误差分布图

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问题，是由于年龄误报问题的存在，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：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和可靠将是一项关系到整个“五普”数据质量乃至中国人口现状与主要问题的研究的重要问题。

从表 10 我们可以看到，调整前 2000 年人口普查时 0－9 岁的人口为 15913 万人，仅仅比国家计生委历年的报表出生人口数多 1201 万人。调整后 2000 年时的 0－9 岁各年龄组人口之和为 20323 万人，比第五次人口普查调查的同批人 15913 万人多出 4411 万人。国家统计局根据历年人口变动抽样调查的出生人口公布的 0－9 岁为 20570 万人。考虑到死亡的因素，我们调整后的结果大致与国家统计局公布数一致。如果按照国家统计局的漏登率计算方法，即漏登率＝漏登人口／实际，那么 2000 年时 0－9 岁的漏登率为 27.7%。

根据这一推算结果，“五普”公报中的漏报人口 2246 万应该是漏报和重报相抵后的净漏报人口。假设漏报全部发生在 0－9 岁人口中，总的漏报人口应该是 4411 万人，而重报人口为 2165 万人。即 126583＝124337＋4411－2165。但实际情况并非如

需要全本请访问 www.ertongbook.com