

1982年中国各民族性别、年龄人口数的推算

黄荣清

【提要】 人口的性别、年龄数,可以说是人口研究中的最基础资料。本文根据中国第三次人口普查(以下简称“三普”)公布的民族人口的各种数据进行整理,在此基础上推算了1982年中国各民族的性别、年龄人口,并以推算的结果为例,分析了一些少数民族人口的现象。

【作者】 黄荣清 首都经济贸易大学人口研究所,研究员。

社会科学研究需要追踪不断变化的社会现象。但当我们试图解释现在时,却发现又不得不回顾过去,在面对那些随着时间的流逝往往变得模糊的历史事件、历史数据而使研究变得困难重重时,我们有时候会情不自禁地责怪前人:为什么当时的学者、当时的研究连一些最基础的研究都没有做?

近年,由中国社会科学院人口研究所发起,各地正在编著各民族人口丛书。在研究该民族人口时,大家自然而然地要用1982年民族人口的性别、年龄资料。因为,人口的性别、年龄是人口学中最基本的内容。它是人口再生产的历史记录,没有准确的性别、年龄人口数,许多最基本的人口指标无从确定,一些人口现象无法解释,高级的人口分析技术无法施展,而进行综合性的研究更无从谈起。但遍查已经出版的“三普”人口资料时,却找不到完整的分民族的人口性别年龄资料。缺了人口的性别、年龄人口资料,许多民族人口研究则变得困难重重,简直难以展开。例如,从1982年到1990年,部分少数民族人口出现高速增长,我们可以肯定自然增加不是主要因素,民族成分的改变是主要原因。但如果我们没有1982年的该民族的性别、年龄人口资料,则难以从定量上加以确定。类似的例子我们还可以举许多。既然分性别、年龄人口数如此重要,如果我们在还有许多线索的时候不去弄清它,那就等于把这难题留给后人,那无异于存心让后人来责难我们。本文意在通过利用现有的人口资料进行推算,补上这一重要的基础数据。

1. “三普”公布的民族人口性别、年龄资料

中国政府历来非常关心和重视民族人口统计工作,为了弄清民族人口的变化,每一次人口调查都把“民族”的调查作为必选项目。例如,中华人民共和国成立后的第一次人口普查,只调查6个项目,其中就有“民族”一项。由于时代的局限,第一次人口普查(1953年)和第二次人口普查(1964年)在汇总调查结果时,只汇总了各民族的人口数(1964年汇总了性别人数),这对于把握民族人口的全貌,是远远不够的。第三次人口普查(1982年),是中国首次使用现代化的技术手段对调查结果进行汇总,其内容的详细程度确实大有进步,是前两次普查不可比拟的。但也正是由于是初次,缺少经验,所以在调查内容和调查的结果的汇总上存在着一些不如人意

的地方,其中,少数民族人口的汇总就是一个例子。“三普”的死亡调查表上没有死亡人口的民族成分,所以我们无法知道当时各民族的人口死亡状况,而公布的民族人口资料中只有分散的资料,没有一个完整、准确的各个民族性别、年龄人口数等等。下面,我们把公布的全国 100% 电子计算机汇总结果^①中有关民族人口性别、年龄的资料整理如下:

(1) 汇总表 3,公布了分民族的总人口数和分性别的人口数;

(2) 汇总表 28 为各个少数民族人口年龄状况,公布了各少数民族在 0~14 岁,15~19 岁,20~24 岁,……60~64 岁,65~99 岁,100 岁及以上年龄的人口数,但没有分性别;

(3) 汇总表 29 为各少数民族人口的文化程度,有各少数民族的 6 岁及以上的人口数;

(4) 汇总表 30 为各少数民族的文盲、半文盲的人口数,有各少数民族的 12 岁及以上人口数;

(5) 在汇总表 74 和表 75 中有各少数民族 15~64 岁女性人口数。

对上述汇总表进行整理,我们可以使民族人口的年龄组别划得更细一些,即可以把 0~14 岁细分为 0~5 岁,6~11 岁,12~14 岁,并得到相应的人口数。关于分性别人口数,则只能分为 0~11 岁,12 岁及以上,及 15~64 岁不连续的分组人口数,但由于 15~64 岁的女性在个别民族上人数有误^②,所以在以下的推算中没有全部利用。

由以上的资料,我们知道,“三普”的 100% 计算机汇总结果并没有给我们提供较为完整、规范分组的各民族性别、人口数。

“三普”公布的资料,除了 100% 计算机汇总结果外,还有 1% 户抽样结果。在公布的 1% 抽样资料中,有各个民族的分性别、年龄人口数。对于人口数较少的民族,如 10 万人以下的民族,按人口的 1% 计算,仅 1 000 人左右,若再按年龄、性别分开,一些年龄仅几个人,有些年龄组人口数甚至为 0。可以设想年龄、性别的误差是相当大的。至于 1 万人口以下的民族,大部分年龄组人口数都是零,其抽样结果根本无法使用。即使对人口数较大的民族,其性别、年龄构成也有相当误差。如把 1% 户抽样结果和在 100% 汇总结果中已知的性别、年龄构成相比较,设:

性别比 = $100 \cdot (\text{男性} / \text{女性})$

年龄组构成 = 年龄组人口 / 总人口。

1% 户抽样的性别比与年龄组人口构成平均误差见表 1。

由表 1 可知,把 1% 抽样调查的结果与 100% 的数据相比较,汉族人口与全体少数民族人口的性别比指数绝对误差在 1% 以内,在重组的年龄组(即分为 0~5 岁,6~11 岁,12~14 岁,15~19 岁,……60~64 岁,65~99 岁,100 岁及以上),人口年龄组构成的相对误差也在 1% 以内。对 80 万人口以上的民族,人口数的误差超过 5% 的有哈尼族;性别比指数误差绝对值超过 2 的有白族和土家族;年龄组构成平均相对误差超过 5% 的仅哈尼族。所以如果按上述这些年龄分组,则 1% 抽样的结果除个别民族外,基本上是可用的。但对 80 万人人口以下的民族,则平均的年龄组构成误差基本上在 5% 以上。误差最大的是高山族,为 96.04%。

由于 1% 抽样调查是以户为单位的等距抽样,可以认为被抽到的样本分布是相当均匀的。即样本的民族人口数大致与该民族实际人口数成比例。按照抽样理论,要达到规定的置信度和误差范围,则必须有相应的样本数来保证,对于一个较小的成数(例如单岁的性别、年龄构成指数),如果规定了较小的误差与较大的可信度,所需的样本数往往是很大的。对单岁年龄

^① 国务院人口普查办公室、国家统计局人口统计司:《中国 1982 年人口普查资料》,中国统计出版社,1985 年。

^② 1982 年人口普查时,西藏未进行关于“死亡”、“婚姻”、“生育”项目的调查,所以公布的 100% 计算机汇总的少数民族生育阶段 15~64 岁的妇女人数,藏族、门巴族、珞巴族和实际人数相差甚远。

表 1 1982 年 100% 汇总结果与 1% 户抽样人口、
年龄构成、性别比的比较

	人口数 (万人)	误 差		
		人口(%)	年龄构成(%)	性别比
全 国	100 391.30	0.08	0.14	-0.01
汉 族	93 667.44	0.00	0.13	-0.03
少数民族	6 723.89	1.23	0.45	0.12
蒙古族	341.14	-1.27	1.47	0.67
回 族	722.84	-0.04	1.29	0.55
藏 族	384.79	-0.85	2.41	1.04
维吾尔族	596.35	1.14	1.76	-1.58
苗 族	502.12	0.55	1.23	1.54
彝 族	545.36	0.90	1.60	0.39
壮 族	1 338.31	0.50	0.97	0.52
布依族	211.93	1.55	2.56	1.90
朝鲜族	176.52	-2.61	2.28	0.18
满 族	430.49	-0.16	1.46	-0.30
侗 族	142.64	-0.02	2.47	0.71
瑶 族	141.20	2.43	1.99	-0.50
白 族	113.22	-2.29	3.81	-3.91
土家族	283.68	-2.25	2.00	-2.61
哈尼族	105.88	-6.79	5.60	1.15
哈萨克族	90.75	1.04	3.55	0.50
傣 族	83.95	0.79	3.30	0.05
黎 族	88.71	2.10	3.98	1.63

注:设 1982 年 100% 汇总结果中 x 岁年龄人口数为 P_x ,
1% 户抽样人口 x 岁年龄人口数为 P_x' , 总人口分别为 P 和 P' 。

$$\text{人口数误差} = 100 \left(\frac{\sum P_x'}{\sum P_x} - 1 \right)$$

$$\text{年龄构成误差} = 100 \sum_{j=1}^{15} \left| \frac{\sum_j P_x}{P} - \frac{\sum_j P_x'}{P'} \right|$$

这里 j 表示正文中按 0~5, 6~11, 12~14, 15~19, ……
60~64, 65~99, 100+ 的年龄分组;

$$\text{性别比误差} = 100 \left(\frac{PM'}{PW'} - \frac{PM}{PW} \right)$$

PM 、 PW 和 PM' 、 PW' 分别表示 100% 汇总结果和 1% 户
抽样中男性、女性人口。

人口数要比实际的少;第二,非封闭人口,这里说的非封闭不是通常说的由于人口迁移造成的人口增加与减少,而是由于在一个民族内有许多人改变了原来的民族成分,造成不同民族人口的此涨彼消。例如,从 1982 年到 1990 年,有相当规模的汉族改变了民族成分,所以推算的 1982 年汉族人口要少于 1982 年登记的汉族人口,反之,有相当规模的人在此基础改到蒙古族成分,所以推算 1982 年蒙古族人口要多于 1982 年登记的蒙古族人口,等等。

以上情况说明,除了少数几个无民族成分改变的民族外,用 1990 年的民族人口资料无法正确地估计出 1982 年各民族人口年龄、性别数。

的要求来说,除了汉族和全体少数民族外,每一个少数民族的样本数可以说都达不到所要求的数。

上面是由普查结果直接得到的民族人口的性别、年龄资料情况。另外,我们还可以设想用间接的方法来估算 1982 年各民族的年龄、性别数。由于在 1990 年全国第四次人口普查(以下简称“四普”)时提供了各民族人口的性别、年龄资料,以及人口死亡情况,我们设想可用倒退存活率的方法来估计出 1982 年的人口数。若设 1982 年、1990 年 x 岁人口数分别为 $P_{82}(x)$ 、 $P_{90}(x)$, 1982~1990 年的生命表中 x 岁的静止人口数为 $L(x)$,则有:

$$P_{90}(x+8)/P_{82}(x) = L(x+8)/L(x) = R(x)$$

$$\text{可得 } P_{82}(x) = P_{90}(x)/R(x)$$

由于没有 1982 年各民族的人口生命表,难以估计 1982~1990 年人口死亡情况,我们暂时用由“四普”资料算出的 1989~1990 年各民族人口生命表(黄荣清、刘琰,1995)。估算的 1982 年各民族人口数与性别比,估算的人口数、性别比,在确定的年龄区间上年龄构成指数与 1982 年实际数据的误差见表 2。

由表 2 可知,推算的 1982 年人口数与实际的人口数对于全国来说,相对误差并不那么大,而对单个民族来说有很大出入。造成两者不一致的原因可能有:第一,由于用的是期末的生命表,存活率高估,由于有存活率高估,则以 1990 年人口为基础推算的 1982 年人

2. 性别——年龄人口数的估计方法

在人口学中,我们知道由洛特卡(Lotka)提出的稳定人口理论,有一套年龄构成的推算方法。但是它是建立在一整套严格规定的条件下才能成立。实际的人口变量(生育、死亡、迁移以及在民族人口中的民族成分的改变)很难满足这些规定条件,所以稳定人口的推算方法在这里并不适用。到现在为止,我们还未看到推算人口性别、年龄结构的普遍适用的方法。只能根据已知的条件灵活选择适当的方法。

事实上,普查提供的资料虽然不是规范、完整的性别年龄资料,但它提供了以下资料可作估算的基础:第一,全国分性别、年龄人口数;第二,各民族的性别人口数;第三,不很规范年龄分组的分民族的人口数;第四,汉族和全体少数民族单岁组性别、年龄估计数;第五,80万人口以上的民族的年龄组人口的估计数。根据上述基础资料,本文采用了年龄差别消去法来估计各民族分年龄、性别人口数(详见黄荣清等,1989)。

设我们要估计 $m \times n$ 矩阵中的元素 x_{ij} (>0) ($i=1,2,\dots,m; j=1,2,\dots,n$); 其中行和列元素之和 $X_i = \sum_j x_{ij}$, $X_j = \sum_i x_{ij}$ 为已知。假设我们先估计一个初值。估计值的行和列元素之和为 $X'_i = \sum_j x'_{ij}$, $X'_j = \sum_i x'_{ij}$ 。

我们要通过调整矩阵元素的值,使它的行和列元素的和和已知的和 X_i 和 X_j 非常接近,并以此作为矩阵元素的估计值;用数学语言来说,若预先规定一个任意小的正数 ϵ , 当有:

$$\begin{aligned} & \sum_i \left| \sum_j x'_{ij} - \sum_j x_{ij} \right| + \sum_j \left| \sum_i x'_{ij} - \sum_i x_{ij} \right| \\ &= \sum_i |X'_i - X_i| + \sum_j |X'_j - X_j| < \epsilon \quad \text{时} \end{aligned} \quad (1)$$

就把 x'_{ij} 作为 x_{ij} 的估计值。

调整方法是这样的:先看一下估计的初值是否满足条件(1),如果满足,就以它作为估计值;如果不满足,则从行(或列)调整开始,令:

$$x'_{ij}(X_i/X'_i) \rightarrow x'_{ij}$$

调整后的行元素 x'_{ij} 之和 X'_i 等于 X_i , 但列元素之和 X'_j 不一定等于 X_j ; 如把调整后的 x'_{ij} 作为 x_{ij} 估计值,看是否满足条件(1),若满足,则停止调整;若不满足,则再调整,这次从列开始调

表 2 根据 1989~1990 年生命表推算的 1982 年人口与实际人口的误差

	人口(%)	年龄构成(%)	性别比
全 国	0.43	1.76	-0.60
汉 族	1.58	1.65	-0.51
蒙古族	-16.23	4.20	1.00
回 族	-1.54	2.54	-0.11
藏 族	-1.87	4.12	-0.85
维吾尔族	2.75	7.21	-1.23
苗 族	-25.17	2.56	-3.13
彝 族	-3.90	2.35	-1.82
壮 族	-0.09	2.55	-1.81
布依族	-4.36	3.42	-1.69
朝鲜族	1.07	2.12	-0.73
满 族	-95.28	9.02	3.22
侗 族	-50.22	4.70	-3.81
瑶 族	-26.81	4.15	-4.55
白 族	-22.20	3.25	-2.54
土家族	-76.80	4.25	-2.66
哈尼族	-0.87	1.19	-1.48
哈萨克族	2.23	3.67	-0.40
傣 族	-3.09	3.27	-0.74
黎 族	-2.89	3.88	-2.08

注:人口、年龄构成、性别比误差的定义同表 1。作为标准的人口为 1982 年 100% 汇总的人口数;比较的人口为推算的人口。

整,令:

$$x'_{ij}(X_j/X'_j) \rightarrow x'_{ij}$$

这时必定有 $X'_j = X_j$;再观察调整后的元素,看是否满足条件(1)。若不满足,则再从行开始调整,可以证明,每调整一次,误差 $= \sum_i |X'_i - X_i| + \sum_j |X'_j - X_j|$ 总有所减小,这样总能得到满足条件的 x'_{ij} 。

这里要特别指出,虽然我们能控制 X_i 和 X_j 的误差,但这样估计出的 x'_{ij} 并不是实际的 x_{ij} 。 x'_{ij} 和 x_{ij} 之间有误差存在。为了控制误差,初值估计是很重要的,我们要尽可能地选择和真值接近的值作为初值。

各民族的人口性别、年龄人口数的估计是分阶段进行的:

第一阶段以全国人口为基准先估计汉族和全体少数民族;第二阶段以估计出的全体少数民族人口为基准,估计 80 万人口以上的每个民族和全部 80 万人口以下的民族人口的和,第三阶段以估计出的 80 万人口以下的民族人口的和为基准,估计 80 万人口以下的各个民族。

估计所以要分成上述三个阶段,一是出于计算机容量的考虑。因为同时对 56 个民族进行性别、年龄人口数叠代所占计算机内存太大,各民族人口数差别大,影响的权数也不同。更主要的是初值条件不同。在第一阶段,以全国人口的性别、年龄人口为基准,对汉族和全部少数民族,可直接用 1% 户人口抽样结果作为各性别、年龄人数估计的初值。在第二阶段,在规定的年龄组,可用 1% 户人口抽样结果,作为 18 个 80 万人口以上及 80 万人口以下的全部少数民族性别、年龄人数估计的初值,但在高龄上作了修匀。在第三阶段,则完全用假设数,先根据已知的男女合计的年龄组人口作单岁插值,估计单岁男女人口,然后用“四普”逆向存活的人口性别比估计相应年龄组性别比,并由此算出男女年龄组人口数作为初值。

在每个阶段,又按两个步骤、两个方向上反复平衡。第一步:估计相应的年龄组性别人口数,在这里,又分为 12 岁以下和 12 岁及以上两个过程;第二步,由估计出的年龄组人口数进行单岁组人口数插值。估计单个民族的性别、年龄人口,然后以基准的人口为标准,分男、女对求出的各民族的性别、年龄人口数进行修正,然后再回到各民族那里去修正。这样通过反复修正,便可估计出每个民族单岁组的性别、年龄人口数。

对用上述方法估计出的每个民族单岁组的性别、年龄人口数,虽然我们不能就此作出结论,它就应该是“三普”时 100% 数据汇总的实际结果,但它至少在以下方面和“三普”100% 汇总结果一致:

- (1) 每个民族总人口、分性别人口数;
- (2) 每个民族 0~5 岁,6~11 岁,12~15 岁,15~19 岁,……65~99 岁,100 岁年龄组的人口数;
- (3) 每个民族 0~11 岁,12 岁及以上的性别人口数;
- (4) 除藏、门巴、珞巴族外,15~64 岁的男女性别人口数;
- (5) 各民族人口数之和等于全国人口数;各民族分年龄、性别人口数之和等于全国分年龄、性别人口数。

由于是在满足上述条件下的推算,我们有理由相信推算出的民族人口的估计值有相当的准确度。

我们知道,关于性别、年龄人口,对人口数较大(如 80 万以上)的民族,现有的系统资料仅有 1% 户的抽样数据,本文的结果也是在此基础上进行推算的。推算的结果肯定比原数据有了

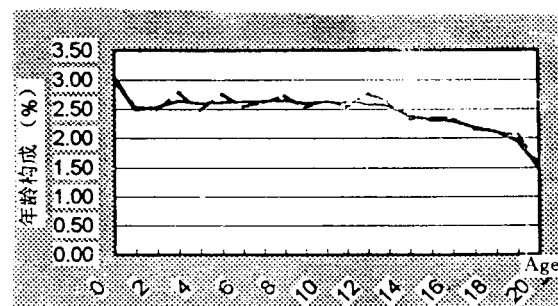
改进。这是因为,除了上述(1)~(5)点以外(原数据不能满足),在以下两点也有了改进:第一,减少了性别年龄人口数的波动。以藏族、哈尼族为例,藏族在年龄申报上,在双数和在以“0”和“5”结尾的年龄上,存在一定程度的堆积(黄荣清,1993),这在1%户的抽样数据中很明显,而在推算值上已消除了这种堆积(见图1);哈尼族的抽样数据不但在总人口性别比上和100%汇总资料不一致,在年龄性别比上也呈高低起伏,肯定有较大的误差,而推算值则有所改进(见图2)。第二,在高龄人口上,直接按抽样数据推算,或者过大,或者过小。例如百岁及以上的人口,在抽样数据中,或者为“0”,即表示该民族无百岁以上的老人,而对于非零的情况,如等于“1”,如按比例推算,则百岁老人又太多了。而本文的推算值克服了这种现象。对于人口较少的民族,1%抽样数据基本上无法利用,是个空白,本文的推算可以说填补了空白。

3. 实例

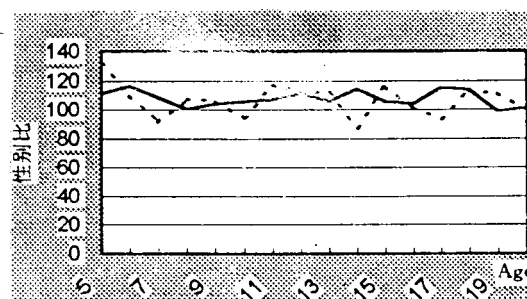
人们常说,人口的性别、年龄数是该人口的历史记录。从性别、年龄人口那里我们可以观察到人口变动的历史轨迹。由于1953年和1964年的人口普查没有对各民族人口进行详细汇总,所以弄清1982年的各民族的年龄、性别人口对民族人口研究就显得格外重要。特别是在1982年,对多数少数民族来说,计划生育政策还刚开始实施,由人口的性别年龄特征可以看出该民族在无人干涉生育下人口再生产的情况;此外大规模改变民族成分的现象尚未出现。从1982年到1990年,大部分民族都不同程度地有原为其他民族(主要是汉族)加入其中(黄荣清,1995),有的甚至占到该民族原人口一半以上。例如土家族,从1982年到1990年,人口增加了一倍多,人口的增加部分,当然主要不是自然增加,而是由原为别的民族后又改变了民族成分加入了土家族而引起的。所以尽管1990年普查也提供了各民族的年龄、性别的详细资料,但由于组成人口的人已发生了很大的变化,由此也带来人口再生产因素(如生育、死亡等)的不同。从这个意义上说,此一人口已不完全是那一人口,要研究民族的历史人口,最好还是用1982年的人口统计数。下面,我们利用已估计出的1982年各民族的性别、年龄人口数,画出一些民族的性别、年龄金字塔,以此可观察到该民族人口的历史变化和再生产的特点(见图3~图8)。我们选择其中较为典型的几个民族,加以说明,希望引起研究者的兴趣。

3.1 全体少数民族

图4显示,中国少数民族人口性别、年龄构成金字塔基本上和全国人口相同,金字塔底宽,但在10岁以下有收缩,收缩的幅度较全国要小。说明从整体上看,少数民族的出生率下降在时间上与全国同步,但平均下降的幅度较全国低。这种金字塔结构代表了大部分少数民族人口性别、年龄的特点。



----- 抽样值; ——— 推算值
图1 年龄构成的比较(藏族)



----- 抽样值; ——— 推算值
图2 年龄别性别比的比较(哈尼族)

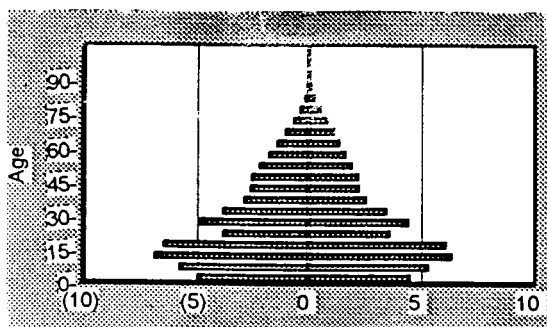


图3 1982年全国人口性别、年龄金字塔

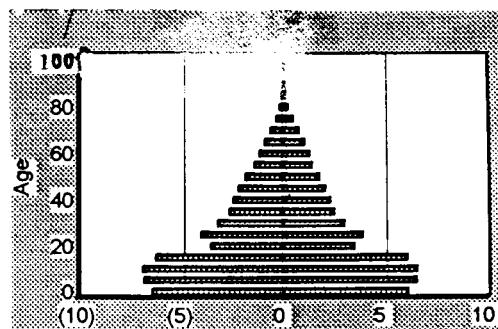


图4 1982年中国少数民族人口性别、年龄金字塔

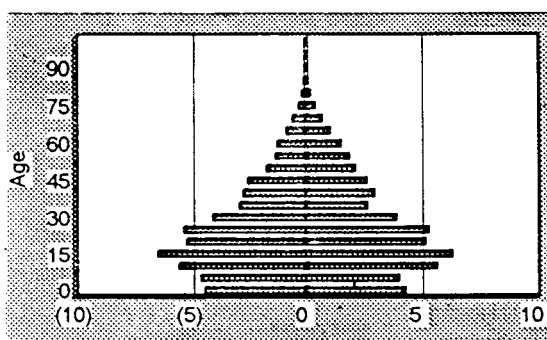


图5 1982年中国朝鲜族人口性别、年龄金字塔

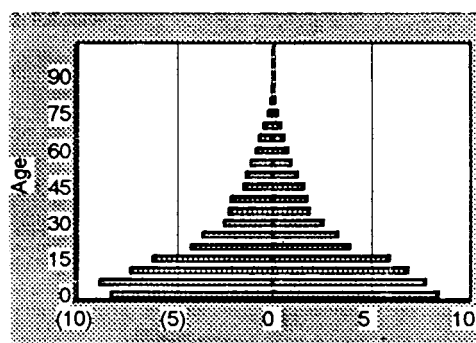


图6 1982年中国哈萨克族人口性别、年龄金字塔

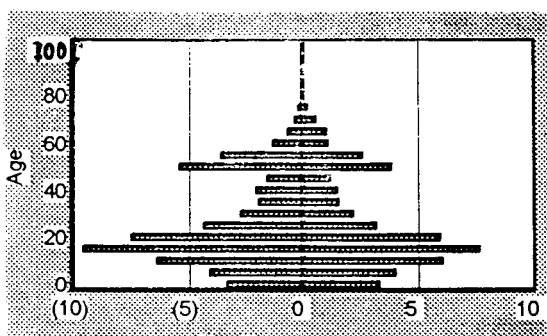


图7 1982年中国大陆高山族人口的性别、年龄金字塔

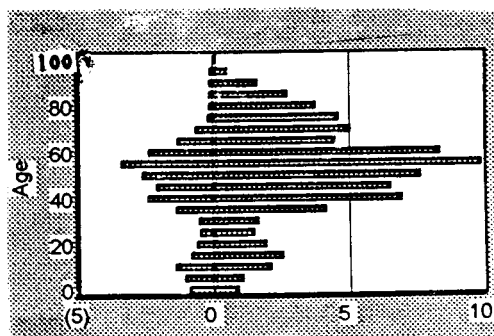


图8 1982年外国人加入中国籍的人口性别、年龄金字塔

3.2 朝鲜族

图5显示,朝鲜族人口性别、年龄构成金字塔是少数民族中较为少见的类型。全国的人口性别、年龄构成金字塔在10岁以下开始收缩,朝鲜族的金字塔在15岁以下就开始收缩,在1~4岁的构成指数都较全国要低,说明朝鲜族在全国广泛开展计划生育以前(70年代初),出生率已开始下降,并且下降的速度较全国要快。

3.3 哈萨克族

哈萨克族人口性别、年龄构成呈标准的金字塔形(见图6)。塔底很宽,在0~9岁的人口年龄构成指数较全国开展计划生育以前(1953年,1964年)还要大,妇女的生育率非常高(哈萨克

族妇女平均生育子女数在 7 个以上)。人口再生产是典型的高速增长型。在金字塔的塔底(0~4 岁),略微有所收缩。说明该民族计划生育开展得较晚,生育率下降也很小。类似这种人口性别、年龄构成的还有撒拉族、克尔克孜族、维吾尔族等。

3.4 高山族和外国人加入中国籍^①

两者人口都很少。1982 年,大陆高山族人口仅有 1 650 人,外国人加入中国籍也不足 5 000 人。这是两个比较特殊的人口,人口年龄构成的类型非常特殊。

从金字塔图形看,大陆高山族人口年龄构成有两个峰值:在 50~54 岁有一个峰值,在 15~19 岁出现第二个峰值。在 35 岁以上性别比明显的不平衡,男性多于女性;而在 30 岁以后性别比趋于平衡。高山族祖居台湾。在 1982 年时,35 岁以上的人为新中国成立以前就在大陆定居,说明当时在大陆定居的以男性为主。新中国成立后,海峡两岸被阻隔,所以年龄人口出现断层。30 岁以下为新中国成立以后出生的人,相应于 50~54 岁的人口峰值,对应出现在 10~25 岁的年龄高峰,且第二代人较第一代人多得多。出现这种情况可能有以下两种因素:第一,50 岁左右的人,当时的生育率较高;第二,与其他民族通婚,他们出生的子女申报为“高山族”。5 岁以下人口较少,可能由于人口年龄构成波动引起的,另外,高山族人口大都居住在城市,虽然计划生育在政策上对他们的生育限制较为宽松,但与他们的先辈相比较,他们也不愿再多生育,所以出现 0~5 岁又开始收缩的现象(见图 7)。

关于侨居外国华人的研究有不少,但对侨居中国的外国人,相应的研究却较少看到。观察 1982 年外国人已加入中国籍的性别、年龄构成,我们可以发现一些有趣的现象(见图 8)。从 1982 年外国人加入中国籍的人口年龄分布看,35 岁及以上的人占绝大部分,约占全部入籍人口的 83%。这部分人又以女性为主,占 80%以上。说明当时已入籍的外国人大部分在新中国成立以前就在中国定居。他(她)们中有不少是抗日战争胜利后遗留在日本的孤儿,且以女性为主。

入籍外国人的第二代要比上一代人少得多。这可能是因为他们和中国人结婚后所生的子女都随中国的父亲(母亲)申报了某一民族,所以这一类人口越来越少。

参 考 文 献

1. 黄荣清等编著:《人口分析技术》,北京经济学院出版社,1989 年。
2. 黄荣清:《中国百万以上人口的民族年龄申报的准确性》,《中国人口科学》,1993 年第 5 期。
3. 黄荣清:《中国各族人口的增长》,北京经济学院出版社,1995 年。
4. 黄荣清、刘琰编著:《中国人口死亡数据集》,中国人口出版社,1995 年。

(本文责任编辑: 朱 萍)

^① 严格地说,“外国人加入中国籍”并不能称之为一个民族。但在中国的民族人口统计中常把他和民族列成一类,所以这里也把他看成类似于民族人口来处理。