

人口年龄结构间接估计方法与应用研究

王 广 州

【提要】 本文全面深入讨论了人口年龄结构间接估计方法,并以人口普查数据为例探讨年龄结构间接估计的精度和相对误差问题,同时指出今后有待深入研究的问题。

【作者】 王广州 哈尔滨工业大学管理学院,博士后。

众所周知,任何人口普查和大规模人口调查的投入都是惊人的。因受人力、物力和财力等因素的限制大规模人口数据收集不可能频繁地进行。因此,在人口数据的开发利用过程中,不能仅仅停留在对现有数据的表面的分析和利用上,应该通过数据内在联系,运用有效的分析方法和分析技术,建立起可靠的分析机制与分析模式,从而达到对现有数据进行深入开发和利用的目的。

无论人口抽样调查还是人口普查都是对人口状态的描述,是时点数据。对于人口研究者来说,仅仅了解、分析人口状态是远远不够的,需要在人口状态描述的基础上,对人口过程进行深入分析。毫无疑问,只有运用比较完善的分析手段,才能将人口状态与人口过程有机地结合起来,进一步对人口系统的行为和功能予以深入的研究。

从人口状态到人口发展过程分析可分为两种情形。一种是对未来人口过程的描述,另一种是对人口发展历史的回溯或重建。对未来人口发展过程的描述属于人口预测问题,其分析方法和分析技术比较成熟,本文不再赘述。对人口发展历史过程的研究则是通过现有人口状态数据对人口发展历史过程的描述,是对人口过程进行回顾性研究。本文将讨论人口年龄结构间接估计方法,试图为探讨如何通过人口状态数据来描述人口系统的发展过程与发展历史创造条件。

一、方 法

人口年龄结构是人口系统基本特征的重要组成部分,是研究人口过程的根本点和出发点。对人口年龄结构的间接估计不仅有利于人口系统的重建,而且对深入研究年龄别人口迁移、年龄别人口城市化过程、人口计划生育效果评估和人口数据质量评估等方面的问题具有重要意义。与此同时,对年龄结构的间接估计还可以解决由于人口普查时点变化所带来的数据比较和数据调整问题。

(一) 利用一个时点人口数据回溯人口年龄结构方法

由于可以获得的人口数据可能是一次调查的数据,也可能是两次调查的数据。利用一次人口调查数据对以往人口年龄结构的间接估计需要采取存活倒推法进行。存活倒推法的基本思想是以生命表为基础,根据年龄别人口存活概率和现有人口结构对历史人口进行推测。具体算法是:建立生命表对人口回溯的起始年份人口年龄别存活概率予以估计(曾毅,1993;Chin Long Chiang,1984;Krishan Namboodiri, C. M. Suchindran,1987)。对生命表中的 a_x 参数本文采用 Coale-Demeny (西区模式)参数。然后,通过回溯方法估算各年份年龄别人口数。年龄别人口数回溯方法的算法为:

$$nPt_1(x) = nPt_2(x+n) * [nL(x)/nL(x+n)]$$

式中 $nPt_1(x)$ 是在 t_1 时刻年龄在 x 岁至 $x+n$ 岁的人口数; $nPt_2(x+n)$ 是在 t_2 时刻年龄在 $x+n$ 岁至 $x+2n$ 岁的人口数; $nL(x)$ 为确切年龄在 x 至 $x+n$ 队列存活人年数; $nL(x+n)$ 为确切年龄在 $x+n$ 至 $x+2n$ 队列存活人年数。

当然,存活倒推法作为人口研究的重要手段不仅可用于对人口过程的回溯,还可与人口预测等方法结合进行计划生育实施效果的评估。

虽然存活倒推法可以用于对人口过程的回溯、重建和人口数据质量评估等方面的应用研究,尤其是该方法成为仅获得一个时点人口数据情况下进行人口间接估计的重要手段。但由于现有数据和存活倒推法本身的问题,存活倒推法对老年人口估计具有明显的不足。其缺陷表现在使用存活倒推法进行人口年龄结构回溯时,每向前倒推一年,人口结构中将有有一个年龄组人口数的估计值无法获得,且倒推的年份每增加一年,无法估计的年龄组也相应地增加一组,即可估计的年龄组减少一组,因此,使用存活倒推法将无法得到完整的年龄结构。另外,由于存活倒推法所采用的死亡数据与历史状况不符,将低估历史年龄别死亡率即高估存活率,因此,如果人口数据准确,那么回溯的年龄别人口数量将低于相应的年龄别实际人口数量。

(二) 利用两个时点人口数据重建人口年龄结构方法

由于使用一个时点人口数据回溯年龄结构的方法存在无法直接重建完整年龄结构的缺陷,因此,为了重建完整的人口年龄结构,需要采用两个时点的数据配合完成。使用两个时点数据的目的在于通过两个时点数据可以弥补使用单一时点数据所造成的人口年龄结构中一些年龄组无法估计的问题,即采用两个时点的数据通过构造年龄别转移矩阵(王广州,2000)与存活倒推法相结合将缺失的数据予以补充,从而避免采用单一时点和单一方法所产生的缺陷。用两个时点数据进行人口年龄结构的重建既可以从起始时点数据出发,运用年龄别转移矩阵方法估计时点间的年龄结构,也可以从终止时点数据出发,运用存活倒推法估计时点间的人口年龄结构。无论采用哪种方法进行估计都存在年龄结构中个别年龄组数据缺失问题。年龄别转移矩阵方法将造成低年龄组人口缺失,存活倒推法则造成高年龄组人口缺失,因此,需要将两种方法结合使用。若以存活倒推法为主要方法重建人口年龄结构,则需要用年龄别转移矩阵方法对缺失年龄组进行补充。类似的,若以年龄别转移矩阵方法为主要方法重建人口年龄结构,则需要用存活倒推法对缺失年龄组进行补充。具体算法为:第一,建立两个生命表,即对人口重建的起始时点和终止时点人口年龄别存活概率予以估计。第二,对时点间其他各时点的生命表的构造方法是采用起始时点和终止时点生命表进行线性插值。第三,构造存活转移矩阵。年龄别人口存活转移矩阵的具体构造方法是: $nPt_2(x+n) = nPt_1(x) * [nL(x+n)/nL(x)]$;式中各变量的具体含义与年龄别人口数回溯方法的算法公式相同。第四,通过回溯方法估算时点间不同时点各年龄组的人口数。

从上述人口重建方法可以看到,对重建的人口年龄结构可以划分为低年龄组人口,高年龄组人口和中间年龄组人口三个部分。低年龄组人口是采用年龄别人口存活转移矩阵方法重建人口年龄结构时无法估计的年龄组;高年龄组人口则是采用存活倒推法重建人口年龄结构时无法估计的年龄组。而中间人口年龄组的重建具有特殊性,既可以用存活倒推法进行估计,也可以用年龄别人口存活转移矩阵方法获得。在具体人口年龄结构重建过程中,采用哪种方法重建中间年龄组人口取决于哪个时点的数据质量更高。如果起始时点的人口数据质量高,就采用年龄别人口存活转移矩阵方法估计中间各年龄组

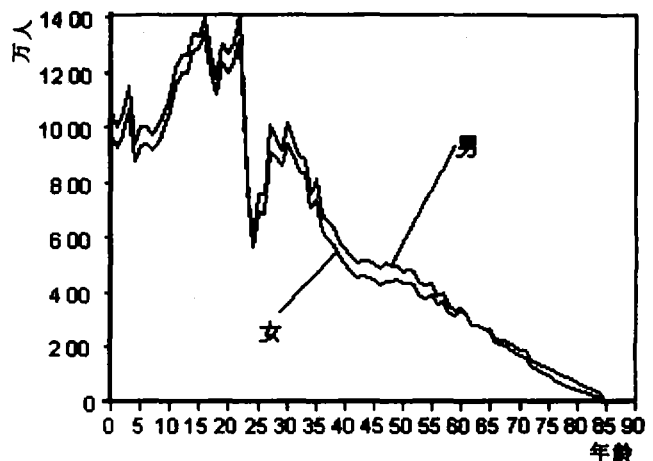


图1 1985年中国人口年龄结构回溯图

人口数量;如果终止时点的人口数据质量高,则采用存活倒推法估计中间各年龄组的人口数量。然而,对于低年龄组和高年龄组人口结构重建的方法则别无选择,即对低年龄组采用存活倒推法进行估计,对高年龄组采用年龄别人口存活转移矩阵方法进行推算。

笔者以 1982 年和 1990 年全国人口普查数据为例,对普查期间的全国人口年龄结构进行回溯、重建,并以此为例来分析和讨论人口年龄结构间接估计的精度和存在的问题。

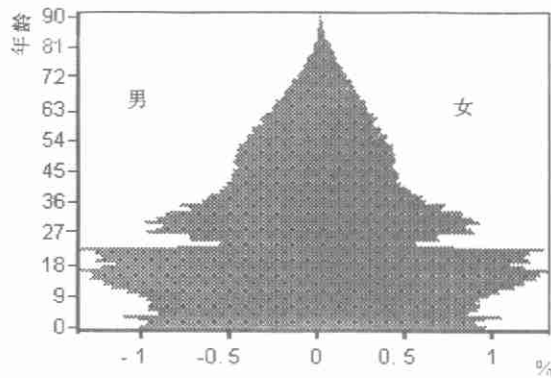


图 2 1985 年中国人口年龄结构估计图

二、应用实例及精度分析

(一) 人口年龄结构估计应用实例

通过 1990 年全国人口普查数据,可以回溯 1990 年以前的人口年龄结构,如可以对 1985 年人口年龄结构进行回溯(见图 1)。

如前所述,由于对人口结构的回溯得到的人口年龄结构是不完整的。为了完整地描述人口普查期间人口系统的发展过程,就需要采用人口重建的方法予以估计。下面利用 1982 年和 1990 年全国人口普查数据对 1982~1990 年间各年的人口年龄结构进行重建,并得到完整的人口年龄结构“金字塔”,从而达到描述完整的人口过程的目的。仍以 1985 年为例,可以采用 1982 年和 1990 年全国人口普查数据重建 1985 年的人口年龄结构(见图 2)。

总之,通过对人口年龄结构的间接估计方法和实例可以看到,利用人口回溯和人口重建方法可以对人口普查间的年龄结构进行完整或不完整间接估计。

(二) 人口年龄结构估计精度分析

从上述实例可以看到,虽然运用人口回溯和人口重建可以对人口年龄结构进行间接估计,但毫无疑问,无论对人口年龄结构回溯还是对人口年龄结构的重建,其精度首先取决于人口数据的质量。数据质量的高低是人口年龄结构间接估计的前提条件,然而,在数据质量一定的条件下,从人口回溯算法的性质可以推断,回溯的时间越长,数据缺失的年龄组越多。同时,由于死亡率的年度差异,回溯的时间越长,得到年龄结构的人口数据与实际数量差距越大。如果死亡数据非常准确,那么,由于过去的年龄别死亡率一定高于当前相应年龄别死亡率,因此回溯的年龄别人口数将低于相应年龄组的实际人口数量。虽然运用两个时点的数据,可以弥补人口年龄结构回溯的缺陷,也就是说,一方面可以获得完整的人口年龄结构,另一方面可以对死亡变化的动态趋势和过程予以逼近,但采用哪一时点并确定相应的主要估计方法却是需要结合具体数据质量状况做进一步研究的问题。

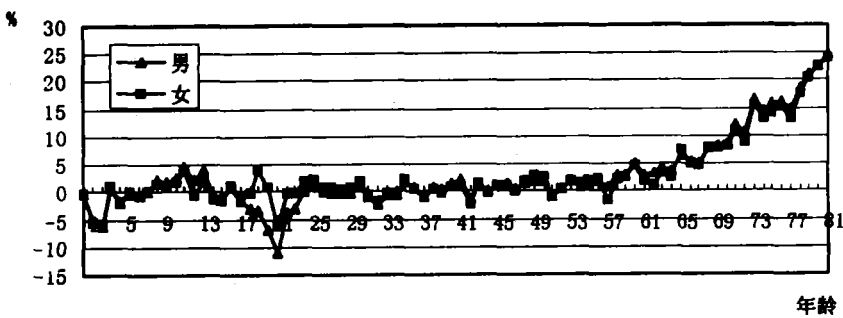
为了用人口回溯与人口重建的方法对人口年龄结构估计的精度进行分析和比较,下面分别以 1982 年和 1990 年全国人口普查得到的年龄结构为评价依据对上述两种方法进行人口年龄结构估计的精度加以比较。相对误差大体分布状况如表所示。

假定人口普查实际收集的数据比较准确,那么估计值的误差应该完全取决于估计的方法。也就是说,年龄结构间接估计的误差主要由人口的年龄别存活概率来决定,因此,随着年龄的增长,由于积累误差越来越大,估计值的误差也应该越来越大。然而,事实并非完全如此。从表可以看到,误差较大的年龄组主要集中在老年人口。因此,无论采用哪种方法估计,老年人口各年龄组人口数量的

表 人口年龄结构估计相对误差比较

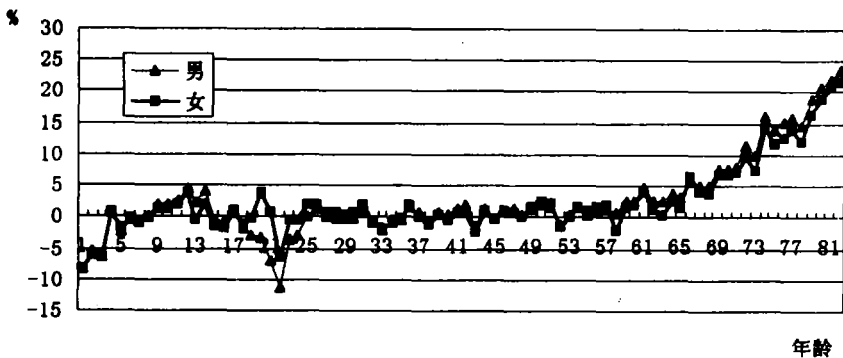
年份	估计方法	年龄组 (岁)	性别	相对误差		平均相对误差	相对误差标准差
				最小值	最大值		
1982	人口年龄结构回溯	0~14	男	0.2780	7.4930	2.7315	2.2365485
			女	0.0062	7.4567	2.3088	2.2793905
		15~64	男	0.0183	11.1685	1.8994	1.9436851
			女	0.0851	6.3009	1.5302	1.2215028
		65~82	男	5.2033	24.7252	13.5278	6.1163874
			女	4.5403	24.0023	12.7624	5.9649425
	人口年龄结构重建	0~14	男	0.1921	8.2211	2.8325	2.4262398
			女	0.0760	8.1526	2.4276	2.4831217
		15~64	男	0.0131	11.2220	1.7943	1.9209232
			女	0.0258	6.4010	1.3783	1.1523749
		65~82	男	5.0588	23.6327	13.0858	5.9980869
			女	3.7374	21.7550	11.5329	5.6436152
1990	人口年龄结构重建	9~14	男	0.5607	6.0640	2.5688	2.4644246
			女	0.1971	5.9305	2.5216	2.5678978
		15~64	男	0.0131	10.0898	1.6762	1.7823863
			女	0.0258	6.0159	1.3130	1.1203511
		65~90	男	0.6592	30.9464	11.5394	8.9961248
			女	0.3985	27.8031	9.8203	8.1990698

注: 相对误差 = |实际人口数量 - 估计人口数量| / 实际人口数量。



注: 相对误差 = (实际人口数量 - 估计人口数量) / 实际人口数量。

图 3 1982年人口年龄结构回溯相对误差分布



注: 相对误差 = (实际人口数量 - 估计人口数量) / 实际人口数量。

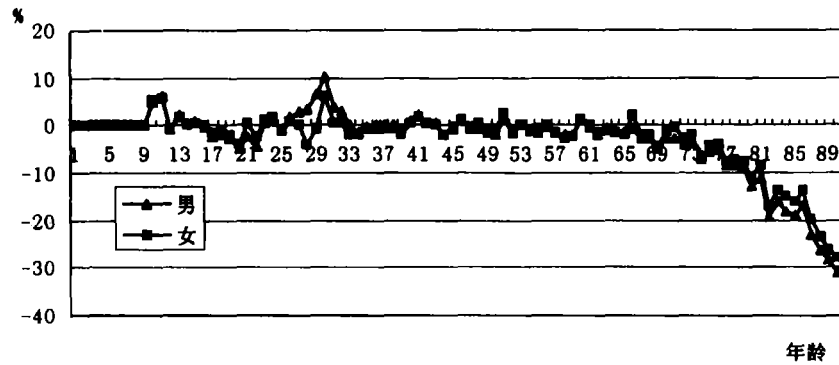
图 4 1982年人口年龄结构重建相对误差分布

估计值的平均相对误差均在 10% 以上。15~64 岁各年龄组的估计值的相对误差最低, 平均相对误差在 2% 以内。0~14 岁各年龄组的平均相对误差则在 3% 以内。用人口重建方法得到的估计值的精度虽然有所提高, 但数量非常有限。其主要原因在于一方面人口死亡相对比较稳定, 另一方面说明老年人口和低年龄组人口死亡数据质量较差。纵观 1982 年和 1990 年人口年龄结构估计值相对误差的分布 (见图 3、图 4 和图 5) 状况可以看到, 男性人口各年龄组的相对误差与相应女性年龄组相差很小, 而且二者相对误差的总体变化趋势完全一致。

(三) 人口年龄结构间接估计的可接受性

从以上对 1982 年和 1990 年人口年龄结构间接估计方法和估计结果可以看到, 不同年龄组相对误差的差别很大 (见图 3、图 4 和图 5)。1982

年全国人口普查登记质量抽样检查结果表明,1982年人口普查年龄差错率为6.15%,死亡人口差错率为4.4%(国务院人口普查办公室、国家统计局人口统计司,1985)。因此,假定1990年全国人口



注:相对误差=(实际人口数量-估计人口数量)/实际人口数量。

图5 1990年人口年龄结构重建相对误差分布

普查数据完全准确,那么,对1982年人口年龄结构估计值的平均相对误差也不可能低于6.15%。然而,1990年全国人口普查登记质量抽样检查结果显示,年龄差错率为3.07%,死亡人口差错率为4.9%(国务院人口普查办公室、国家统计局人口统计司,1993)。同样,无论根据1990年全国人口普查数据估计1982年人口年龄结构,还是根据1982年全国人口普查数据估计1990年全国人口年龄结构的平均相对误差不可能低于3.07%。因此,根据1990年全国人口普查数据推断1982年人口年龄结构的精度不可能非常准确,其相对误差在3.07%以内年龄组的估计值应该是可以接受的;而对于老年人口的估计则因误差较大,有待今后进一步深入研究。由此可见,在重建两次普查间人口年龄结构时,对老年人口的估计应该使用1982年的数据进行。对1990年人口年龄结构误差的调整也应该采用1982年人口数据来实现。

总之,根据上述人口年龄结构间接估计算法和应用实例分析可以得到以下结论:(1)利用一个时点的数据运用“存活倒推法”进行人口数量回溯将低估各年龄组的人口数。(2)利用两个时点的人口数据,对时点间的人口年龄结构进行重建,其重建结果更接近实际情况。(3)在生命表构造过程中,由于数据质量和累计误差等方面的原因,老年人口年龄别存活概率与实际情况的差距较大。相应地,对老年人口的回溯或重建的结果的相对误差较大,其估计值的精度远远低于劳动适龄人口。因此,如何降低老年人口重建的误差将对降低总体估计误差起到重要的作用。(4)由于采用一个时点人口数据进行人口回溯产生高龄人口年龄组数据无法估计的问题,因此,有待今后结合稳定人口或动态人口分析方法等对此进行深入研究。

参考文献:

1. 曾毅:《人口分析方法与应用》,北京大学出版社,1993年。
2. 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(电子计算机汇总)》,中国统计出版社,1985年。
3. 国务院人口普查办公室、国家统计局人口统计司:《中国1990年人口普查资料》,中国统计出版社,1993年。
4. 王广州:《中国人口与市场信息系统的开发研究》(博士论文),北京大学,2000年。
5. Chin Long Chiang(1984), *The Life Table and Its Applications*, Robert E. Krieger Publishing Company.
6. Krishan Namboodiri, C. M. Suchindran(1987), *Life Table Techniques and Their Applications*, Academic Press, INC.

(本文责任编辑:朱 萍)