

人口统计研究中方法的误用与滥用*

——以 P/F 比方法为例

郭志刚

【摘要】人口统计的方法和技术是帮助人们从数量上分析和认识人口现象的有力工具。但是,倘若运用不当,即使是科学的方法和技术,也有可能得出错误的结论,甚至会成为谬误的护身符。近年来,人口研究中误用或滥用统计分析方法的现象经常出现,并已影响到人口统计的严肃性和人口分析的准确性。文章以新近采用 P/F 比方法检验中国 2010 年第六次全国人口普查生育数据的一项研究为例,具体说明如果统计方法使用不当,特别是若无视方法内在包含的前提假定,便会得到误导性的结论。因此,人口研究中应当谨防方法的误用与滥用。

【关键词】人口统计 P/F 比方法 方法适用性

【作者】郭志刚 北京大学中国社会发展研究中心研究员、社会学系教授。

一、问题的缘起

人口统计学一直被认为是擅于量化的一门社会科学学科,建立了相当完善的对生育、死亡、迁移等基本生命事件进行统计分析的方法和技术。人口学专业人才培养中必定需要开设人口统计学课程,对相关人口统计方法和技术进行专门训练。正是借助于这些方法和技术,人口学研究者才能得以对人口现象进行科学严谨的统计分析,揭示其背后的规律,人口学研究才能得以不断繁荣和发展。

从方法论上讲,任何模型、方法和技术都有其适用条件,或者说都包含内在的假定条件,只不过这些假定会存在或强或弱的差别。比如,近年来国内外人口学者关于进度效应对时期总和生育率影响的讨论提供了一个重要启示,就是当一个人口进入低生育率状态时,生育时机的改变会对常规总和生育率计算结果产生不可忽略的影响,从而显著扭曲其作为终身生育水平估计的功能。因此,Bongaarts 等(1998)提出去进度效应总和生育率指标,以避

* 本文为教育部人文社会科学重点研究基地重大项目“中国低生育率研究”(批准号:12JJD840005)的阶段性成果。

免仅看总和生育率指标而得到误导性的认识。实际上,现有人口统计中有不少方法和技术都建立在稳定人口这个基本假定之上,特别是那些被称为间接估计的方法和技术。然而时至今日,很多发达国家,甚至像中国等不少发展中国家,其人口仍处于迅速转变时期,生育水平和存活水平仍在不断变化,明显是一种非稳定人口状态。换言之,这种人口现实与稳定人口在特征和属性上迥异,导致以往很多人口统计方法失去了应用的基础。所以,今天我们对“新常态”下的人口现象进行统计分析时,一个必须加以考虑的问题就是所用的人口模型、方法和技术是否仍合时宜?

人口统计方法虽然是定量分析和认识人口现象的有力工具,但倘若运用不当,即使是科学的方法也有可能得出错误的结论。近年来,人口研究中误用或滥用统计方法的现象屡见不鲜,已经严重损害了人口统计的严肃性和人口分析结论的可信度。为此,本文以陈卫、杨胜慧(2014)用 P/F 比方法检验中国 2010 年第六次全国人口普查(以下简称六普)生育数据的研究为例,具体说明对处于“新常态”下的中国人口现象进行研究时,必须考虑所用技术方法的适用条件,尽量避免统计方法的误用或滥用,尽量避免得出误导性的错误结论。

二、案例文章中的错误

陈卫、杨胜慧(2014)采用 P/F 比方法检验了中国六普的生育数据。该文使用的是国际著名人口学家提出的成熟方法(Brass 等, 1968),而且未借助任何其他来源数据,仅使用“系统内”的 2010 年人口普查数据对生育率结果做了评价,发现六普生育统计的 P/F 比值在 20~24 岁到 35~39 岁之间各年龄组上都保持着较为平稳一致的值,即均在 1.4 左右。该文将这种结果解读为,在过去的 15 年里,中国的生育水平一直变化平稳,没有出现进一步的下降趋势。于是,这个比值 1.4 便可以表明六普收集的时期生育数据有严重漏报,导致其公布的总和生育率低报了 40%。然后,他们再用这个比值作为调整系数对六普生育率结果加以调整,将六普总和生育率从公布的 1.18 提高到 1.66。然而,对这一结论笔者感到不解。笔者承担过国务院人口普查办公室委托的六普生育分析课题(郭志刚,2014),研究结果不仅发现六普总和生育率比五普相应统计(1.22)又略有降低,而且六普数据还能反映出这种变化有着多种社会经济和人口原因。比如,21 世纪前 10 年中,年轻一代的受教育程度有所提高;随着城镇化进程的加快,农村人口结婚和生育年龄明显推迟。城镇人口的婚育年龄早就处于不断推迟之中,2005~2010 年,农村妇女的婚育年龄也出现了显著的变化。六普结果表明,20~35 岁农村未婚女性比例明显提高,尤其是 22~25 岁,未婚比例提高均在 10 个百分点以上。此外,农村育龄妇女进城的比例已非常大。2000 和 2010 年全国人口普查表明,城镇化比例分别为 36.9%和 49.7%。育龄妇女人口城镇化比例分别达到 40.7%和 54.2%。而 20~30 岁生育旺盛期妇女城镇化比例更高,2010 年六普时已达到 56.8%。这些重要的变化不仅导致六普生育水平的降低,也导致生育年龄模式发生重大变化。

P/F 比方法的应用条件不仅要求生育水平稳定,而且要求生育模式稳定。陈卫、杨胜慧

(2014)对此假定条件也有明确交代^①,但该文仅凭六普数据的各年龄组 P/F 比值平稳一致便断定过去 15 年生育率变化平稳,完全没提及文献中关于近年中国生育模式变化的研究结果,自己不去具体分析中国生育模式是否有所变化,也没有认真分析六普的 P/F 比值的平稳会不会出于其他原因,就简单地将该指标结果作为调整漏报影响的系数。该文发表时仅用六普生育数据 P/F 比分析一个孤例,没有将中国 1982、1990 和 2000 年三次全国人口普查(以下简称三普、四普、五普)数据也进行同样的分析。按说有了对六普分析的计算模板,其他三次普查的各相应队列的曾生子女数和时期年龄别生育率统计都是现成的,只要套入计算程序很容易得出结果,但该文却没有进行这些分析。该文中还有其他一些评论或结论也是在论证不足情况下的唐突断言,难以令人信服。比如,该文显然很赞同“没有充分理由认为 2010 年普查开展时的社会环境和普查的登记办法,能够使 2010 年普查质量比 2000 年普查更好”的观点。但笔者觉得这种论证及结论均有些不可思议,因为现实是 2000 年以后社会经济迅速发展,教育水平提高,城镇化推进,因而城乡婚育模式有显著的变化。此外,2005 年 1% 人口抽样调查试行了流动人口在户籍地和现住地双重登记,就是为了减少漏报。而六普延续了这种登记方法,在其他方面也采用很多办法来提高数据收集质量。有些学者居然视而不见,这不能不说是一种成见或偏见。近 20 年来,时期总和生育率不断走低,已远低于更替水平,甚至低于生育政策的要求。越来越多的学者认识到这种低生育水平已经不能简单归结为生育政策限制的结果,在很大程度上是出于社会环境的变化。然而,在各方面都迅速变化的总形势下,一部分人口学者却坚持认为社会环境没有变化,生育水平和生育模式也没有变,那么超生数量和比例也没有变,因而生育或出生的漏报和瞒报的必要性也不会变。但是,笔者认为这种推理逻辑难以成立,其实真正没变的是这些学者观察现实的眼光与思维。

多年来,一些人口学者一看到调查数据与其他登记数据之间存在差别,或者某种调查数据内部之间的统计结果存在差异,不管各个系统的数据收集特点不同、服务目的不同、统计口径不同,便简单归结为调查数据漏报,而且一律按数大为准的标准来评判质量。陈卫、杨胜慧(2014)也是这样,一看到普查数据的低龄妇女组的曾生子女数小于计算相应年龄组生育率时所用的出生数,便认为不合逻辑,立刻视为出生漏报。实际上,只要认真查询便不难发现,普查低龄妇女组的曾生子女数小于前 12 个月内妇女的生育数本是正常的,因为这两种生育数的统计口径不一样。另外,陈卫、杨胜慧(2014)断定,“现在看来,无论使用什么数据和方法,2000 年的生育率^②不会低于 1.5”,并且认为,“在上面所述的 2010 年普

① 该文表述为:“如果生育率(包括生育水平和生育模式)长期稳定不变,那么妇女各个年龄的平均曾生子女数和相应年龄的累计生育率,或终身生育率和总和生育率,应该完全相等。即使生育率发生下降,这两项指标在较年轻年龄组上也应该比较一致。”

② 这里“生育率”应为“总和生育率”。原文中多处将“总和生育率”简称为“生育率”。以下引文同此。

查和2000年普查得到的生育率类似,且从社会环境和普查登记办法看都没有支持2010年普查数据质量比2000年普查更好的证据的情况下,我们也有理由认为2010年普查的生育率不会低于1.5。”然而,该文一开始对中国普查生育率的类型出现概念错判,继而出现计算操作不当,最后又出现方法逻辑上混淆不清,因此其所有计算结果和推论也都失去了合理性。

笔者重新研读了方法创建者的原文献,不仅计算了三普、四普、五普和六普的P/F值,还收集了其他一些人口的生育数据做了试算。尽管这些数据未必完美,试算结果也不一定都可靠,但是,试算可以帮助我们更好地理解P/F比方法背后的理论和逻辑,也可以增加一些实际感悟,再经过分析和归纳,大致可以得出P/F比方法并不适用于六普生育数据的结论。狭义的原因是由于几十年来中国的生育水平和生育模式一直在显著变化。广义的原因是该方法并不适用于生育水平很低的人口。因此,根据P/F比方法得到的六普生育率间接估计并无实际参考价值。

下面笔者将分几个方面来报告自己应用P/F比方法的试算结果和其他相关分析的发现。

三、五普和六普的不同生育统计口径

若以普查生育数据作为分析研究的对象,首先研究者应当做到十分熟悉普查生育数据的统计口径。由于P/F比实际上是将相应年龄组妇女的曾生子女数与累计年龄别生育率相比,所以此类研究需要认真考察在中国人口普查中是如何定义年龄别曾生子女数的,特别是如何计算年龄别生育率的,比如年龄组如何定义,各年龄组的妇女人数和出生人数又如何定义,不同的定义计算出的年龄别生育率统计结果是不一样的。陈卫、杨胜慧(2014)断定,“对于任何年龄组,都应该是曾生子女数大于出生人数,因为出生人数只是普查前一年的出生人数,而曾生子女数确是^①普查前多年来的累积的出生人数。理论上有可能存在这两者相等的情况,但实际上应该是曾生子女数大于出生人数,而绝对不可能出现曾生子女数小于出生人数的情况。”基于这种认识,该文过早地下了结论,认为大多数人口普查或小普查数据中都出现过15~19岁育龄妇女组曾生子女统计数显著小于生育率计算中相应年龄组的出生子女数是不合逻辑的现象,是“绝对不可能出现”的情况。其实,陈卫、杨胜慧(2014)已做出判断,认为这一统计现象是出生漏报导致的不正常的统计结果。

根据笔者多年对人口普查生育数据进行研究的经验,陈卫、杨胜慧(2014)指出的这种不匹配现象确实常常出现,但却是普查统计中的正常现象。这首先涉及普查生育统计的年龄口径问题。人口普查公布曾生子女数和普查前一年生育率时虽然都按年龄组(包括一岁组和五岁组)提供,但必须注意,这两种统计表中年龄的含义实际上并不相同。平均曾生子女数表格中的年龄组是按普查标准时点上妇女确切年龄划分的,这个概念很简单。而生育

^① 这里“确是”疑为原文出现的错别字,应为“却是”。

数据表中的年龄却是时期概念,即在普查前一年中暴露于某一年龄生育风险的育龄妇女人年数,因此这个年龄指的是历险年龄而不是普查时点年龄。熟悉列克西斯图的人都知道,在一年中暴露于某一整数年龄生育风险的其实会涉及两个相邻的单岁队列。因此,中国普查公布的生育率表中往往将这个历险人年数称为平均妇女人数。有的国家则在生育率表格中直接称为暴露数(exposure),其概念表达更为准确。

如果了解这种不同,理解15岁队列的曾生子女数和计算15岁生育率时用的该年龄组妇女出生数之间的差异其实并不难。比如,按国务院第五次全国人口普查办公室下发的《第五次全国人口普查机器汇总表式》中根据第五次人口普查数据对育龄妇女分年龄生育率的计算方法是:

$$x \text{ 岁平均育龄妇女人数} = \frac{x \text{ 岁育龄妇女人数} + (x+1) \text{ 岁育龄妇女人数}}{2} \quad (1)$$

$$x \text{ 岁的出生人数} = \frac{x \text{ 岁有生育的妇女人数} + (x+1) \text{ 岁有生育的妇女人数}}{2} \quad (2)$$

按照公式(2),普查生育率中用的出生人数其实是“平均出生人数”。如果简单假定普查时15岁和16岁的妇女人数一样多,那么,由于普查的生育表中的15岁出生人数其实是普查时点上15岁和16岁两个队列出生人数的平均数,也就是公式2中表示的15岁和16岁有生育的妇女人数的平均数。而16岁妇女生育水平一般比15岁高,所以在假定这两个年龄组妇女人数一样多的条件下,两个年龄组的平均出生数自然会大于15岁妇女调查时的曾生子女数。何况,现实普查中还会出现15岁妇女数少于16岁妇女数。比如,四普和六普时便是这种情况,那么这两种不同口径的年龄别生育数之间的差距还会变得更大。因此,普查时点上15岁妇女的曾生子女数常常会小于生育表中15岁的这个“平均出生人数”。其实,这些都是人口统计学的常识。也就是说,这两个表中的出生数本来反映的就是一回事,所以它们之间数量不等并不值得大惊小怪,更不能简单地视为普查出生漏报的反映。

实际上,在中国人口普查中对于生育率的统计口径的确有过几次变化。比如,笔者在分析六普生育数据时发现,六普公布的长表(10%抽样)生育率曲线上存在一些明显的年龄别波动,其原因是六普的生育率采用了与五普不同的计算方法。笔者对六普供分析用的再抽样样本(相当于总人口1%抽样)分别采用这两种计算口径的测试结果表明,用六普口径计算的年龄别生育率曲线具有与公布数据同样的年龄别波动,而用五普口径计算的年龄别生育率曲线却相当平滑(郭志刚等,2014:33-34)。

笔者曾在给六普招标课题组下发的长表年龄别生育率(L6-05)表格的脚注中找到了六普生育率的计算方法。可以概括为两条:(1)六普生育率计算年龄别平均育龄妇女人数时沿用了五普方式(即常规平均数方法);(2)在计算年龄别出生人数时与五普方式不同,采用了按普查时点前一年时间的期中(即2010年5月1日零时)作为标准时间来划分生育率表中的育龄妇女年龄,并直接汇总出相应年龄组的出生数。

经验表明(郭志刚等,2014),对六普 1‰再抽样数据用不同统计口径计算的生育率模式会存在一定差别,其中六普算法得到的低龄组生育率略低于五普算法的相应结果。但是,两种算法各自汇总的总和生育率差别很小。比如,对六普 1‰样本用五普方法计算的总和生育率为 1.190,而按六普算法的统计结果则为 1.188,且后者恰好等于公布的长表生育率结果。其实,六普算法的总和生育率略低是与其低龄组生育率略低紧密相连的,而这一特点对于本文后面的 P/F 比方法探讨十分重要。然而,这两种不同计算方法得到的年龄别生育率则会对 P/F 比结果产生非常大的影响,其中最重要的是这两种生育率结果在应用P/F 比方法时应该分别套用两组不同的系数。

实际上,P/F 比方法的计算公式是固定的,但有两组计算系数分别供不同生育率口径使用。其中,a 组系数适用于调查前一年的出生是按母亲调查期末年龄分组的生育率口径;而 b 组系数则适用于前一年中的出生是按母亲生育时年龄分组的生育率口径。陈卫、杨胜慧(2014)提供的表 1 实际上便是其中的 b 组系数,他们不太清楚普查时点年龄和生育事件的历险年龄之间的差别,想当然地认为六普生育率中的分子(即出生)自然应该是按母亲生育年龄划分的,因此在计算六普的 P/F 比时用 b 组系数,只是在脚注里提到还有一套 a 组系数。

对比五普和六普的生育率口径。其实五普是将两个单岁队列妇女人数平均的方法估计了一年中的历险人数,并用同样方式估计了相应出生人数,因而更接近于 b 组系数对应的生育率口径。但在六普时,生育率历险人数的统计方法虽然与五普相同,出生事件数其实是按调查期末妇女年龄划分的,只不过在名义上将其提前半岁。换句话说,六普生育率表中 15 岁组出生数其实是普查时点上[15.5, 16.5)岁的妇女所生^①,所以该队列妇女在普查前 12 个月内经历了 3 个年龄的生育风险,即 14 岁、15 岁和 16 岁(见图)。这样定义的出生仅仅来自于一个单岁妇女队列,而理论上的 15 岁生育率的分子则应该涉及两个单岁妇女队列在调查前 12 个月内于 15 岁上的出生。如图所示,如果按 2010 年 5 月 1 日时妇女年龄来汇总 15 岁及以上的出生,那么普查时点上年龄在[15.0, 15.5)岁之间的妇女所生的子女数就会从生育率统计中被删

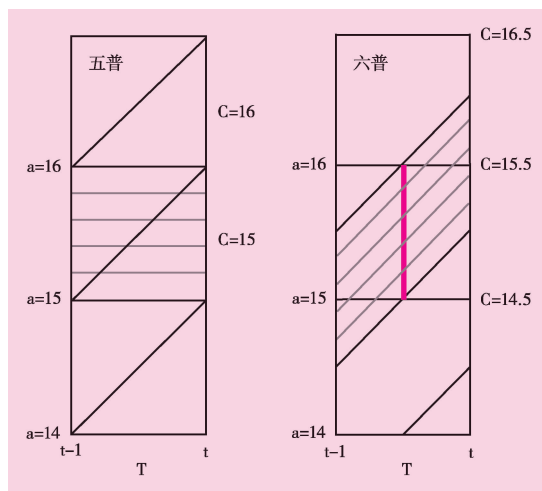


图 五普和六普对年龄别生育率出生数的统计口径图示

注:a 为年龄,T 为普查时点的前 12 个月内,C 为按普查时点年龄为下限的单岁妇女队列。图中用阴影区域来表示出生定义口径涉及的风险区域。

① 六普时[15.0, 15.5)岁妇女的生育是如何处理的也并不清楚。

截掉^①。于是,与五普公布的15岁组生育率中出生数口径相比,六普相应统计数能否与之对应的条件为,图中这个C=15.5岁的单岁队列于前12个月内在14岁与16岁的合计出生数是否等于C=14.5岁队列同期中在15岁出生和C=16.5岁队列同期中在15岁出生的合计数。但这是很难相等的,因为队列妇女数差异、生育水平上的年龄差异和队列差异都会对此有影响。

然而,以上探讨对P/F比方法最重要的意义是,六普的出生不是按母亲生育时年龄来划分的,而是按普查时点年龄来划分的,因此陈卫、杨胜慧(2014)对六普出生采用b组系数计算P/F比值是不恰当的,应该采用a组系数来计算。

四、试算历次人口普查数据的P/F比

中国从1982年第三次全国人口普查起便提供详细的各年龄组育龄妇女的平均曾生子女数,把这些数据与各年龄组的时期生育率结合起来,便可以计算P/F比值。这些普查生育数据并不难获得,但在应用P/F比方法时存在一些数据口径问题。

首先,三普调查的标准时点为1982年7月1日,而收集时期生育数据的时间口径却是1981年这个日历年,因此出生数和妇女数两者在时间上相差半年^②。1990年第四次全国人口普查对此问题进行了修改,其标准时点为1990年7月1日,收集了1989年的上半年、下半年和1990年上半年的生育情况数据,但公布时仍是类似三普的1989年日历年生育率。三普和四普收集生育数据时只需要育龄妇女回答规定时期中是否有过生育,并不询问妇女的具体生育时间,这是与五普和六普的不同之处,所以,三普和四普数据根本无法确定妇女生育时的确切年龄。四普曾生子女数表的 $x+1$ 岁妇女数与生育数据表中 x 岁妇女数完全相等,说明四普时是按调查时妇女人数的年龄减1岁直接作为生育数据表中的妇女人数的,那么三普应该也应用了这种简单汇总方式。因此,三普和四普生育数据可以视为对应P/F比的a类系数,即按队列回溯母亲人数及其出生人数,所以尽管各队列在普查时有相应年龄,然而,严格地说普查提供的是时期队列别生育率,而不是真正的时期年龄别生育率。

五普和六普都询问了妇女的生育月份,因此本来可以按妇女生育时的年龄来汇总时期年龄别出生。但如前所述,五普是用两个队列妇女人数和出生数平均的方法估计前12个月内的生育历险人数及其相应出生人数,因此是时期年龄别生育率的一种简单估计。而六普汇总方式则介乎于四普方式和五普方式之间,妇女人数沿用了五普口径,是时期年龄别口径,但出生数则类似于四普的时期队列别口径。

总之,为了对P/F比进行探测,我们权且忽略普查数据中的其他口径问题,分别按a组系数和b组系数计算各次普查生育数据的P/F比值,结果如表1所示。

① 六普生育率统计中将普查时点上高端年龄[50.0, 50.5)岁的妇女生育包括在内。相应的普查问卷中询问前12个月内生育时要求50周岁的妇女也要填报。

② 查瑞传、季咏华(1984:618)在对三普生育的分析中曾明确指出过生育率的这个问题。

表 1 显示,无论是用 a 组系数还是用 b 组系数计算,在 15~19 岁低龄组中,三普和五普的 P/F 比都出现了异常的负值,而四普低龄组的比值为正值。三普、四普和五普用两组系数计算的 P/F 比值在 20~39 岁各年龄组中同样呈现出随年龄而明显上升的现象,反映出生育水平或生育模式在变化。在这种情况下,年龄别 P/F 比值变化不能简单视为数据质量有问题,P/F 比值水平(如 20~39 岁比值的平均值)也不能作为生育率间接估计的调整系数。值得注意的是,如果忽略年龄别 P/F 比值的具体变化,这三次普查 20~39 岁 P/F 比平均值反映出一种不可思议的结果,即“三普生育率低报了 24%~34%,四普生育率略有高报,而五普生育率则比较准”。这一结论显然与学者们历来普遍认为三普质量最好、五普质量极差的观点相反。总之,用两组系数试算这三次普查的 P/F 比值虽然水平有差异,但年龄别动态却差不多。

表 1 中国四次人口普查的 P/F 比值的试算结果

年龄组(岁)	三普	四普	五普	六普
按 a 组系数				
15~19	-1.9171	0.8879	-2.8442	2.6588
20~24	1.1189	0.7819	0.7130	1.2023
25~29	1.0541	0.9186	0.9509	1.3188
30~34	1.2448	0.9683	1.1398	1.3637
35~39	1.5395	1.1119	1.2690	1.4012
40~44	1.8040	1.4114	1.3953	1.4836
45~49	2.0589	1.7462	1.5118	1.5626
20~39 岁均值	1.2393	0.9452	1.0180	1.3210
按 b 组系数				
15~19	-0.7644	3.2486	-0.9093	-45.4050
20~24	1.3777	0.9090	0.8411	1.4254
25~29	1.1448	0.9663	1.0008	1.4123
30~34	1.2696	0.9818	1.1547	1.3977
35~39	1.5501	1.1169	1.2723	1.4137
40~44	1.8095	1.4135	1.3964	1.4909
45~49	2.0599	1.7467	1.5121	1.5646
20~39 岁均值	1.3355	0.9935	1.0670	1.4120

表 1 中按六普 b 组系数计算的结果与陈卫、杨胜慧(2014)的结果完全相同。六普的 15~19 岁组的 P/F 比值出现了异常的负值,而在 20~39 岁区间各组的 P/F 比值基本处于 1.4 左右。但是,换用 a 组系数计算六普的 15~19 岁组的 P/F 比值却是符合逻辑的正数,更重要的是其他各年龄别 P/F 比值都是随年龄而升高的,于是 b 组系数那种 P/F 比值“平稳”结果便不复存在。a 组系数的 P/F 比值反映以往生育水平或模式有显著变化,尽管 a 组系数计算六普 20~39 岁组 P/F 比平均值也高达 1.32,但这一结果不能视为六普生育漏报的证据,自然也不能用于调整六普总和生育率。

P/F 比方法应用的必要条件是生育水平和模式长期稳定,或者说更适用于一个稳定人口,那么各年龄组妇女人数均应保持稳定比例,那么真正决定年龄别生育率水平的便是年龄别出生人数。从这个意义上讲,笔者针对六普生育率的实际口径采用 a 组系数计算的 P/F 比值,应该更符合该方法的要求。然而,与其他普查不同的是,用 a 组系数居然结果大相径庭。因此,陈卫、杨胜慧(2014)的一系列主要结论完全被颠覆。至于该文发现其 P/F 比所取得的生育率估计值与他们先前使用普查数据和公安数据对 2000 年以来的生育率估计值很一

致(陈卫、杨胜慧,2014),到底是一种“偶然的巧合”还是存在某种共性原因,笔者无从评价。但是,a组系数的计算结果实际上与普查公布的生育率、年龄模式,以及其他多项全国调查反映的生育结果是一致的,而且这些结果均是基于调查原始数据的直接分析。

除了六普应用b组系数的结果以外,表1中所有其他两套系数的试算结果都表现出P/F比值随年龄而升高的特征,反映出中国几十年来生育水平和生育模式一直处于变化之中。陈卫、杨胜慧(2014)认为,“虽然Brass提出的这种P/F比值方法被应用于非洲国家及生育率较高且相对稳定的国家,但中国生育率在过去十多年里较为平稳,也满足该方法的适用条件。”这一判断建立在不恰当地应用b组参数计算P/F比的结果之上,而在此之前该文并没有对实际生育率进程是否平稳本身加以认真研究。因此,这一判断是在用间接估计结果来反证实际动态本身,属于基于假定的间接估计结果来反证这些估计的前提假定合理,无异于同义反复,违背了科学研究逻辑。换句话说,年龄别P/F比值不平稳大致可以反映生育情况有变化,但年龄别P/F比值平稳却并不一定能证明实际生育情况平稳,因为它只是一个必要条件,不是充分条件。

五、六普1‰样本数据的P/F比

对六普公布生育率分别应用两套不同组系数时会有出入很大的P/F比结果,这提示我们应深入探讨P/F比方法的性质。在笔者承接六普委托课题时,取得了六普1‰数据样本按母亲的普查年龄(即六普口径)和按生育年龄(即五普口径)分别汇总的两套年龄别生育率,在这些不同口径生育率统计的基础上,可以应用P/F比方法对六普情况再多做一些测试计算。表2中提供了对六普1‰样本按生育率口径和按P/F比两组系数交互搭配的测试结果。

表2显示:(1)无论是按母亲生育时年龄(即五普生育率口径)还是按母亲在普查时年龄(即六普生育率口径)计算的生育率,采用a组系数得到的P/F比值水平都相对较低,而采用b组系数时则水平相对较高。(2)“六普口径—普查年龄a”和“五普口径—生育年龄b”都是生育率口径正确对应系数分组的结果。一方面,我们发现这两种“正确”组合取得的20~39岁P/F比均值分别为1.2527和1.3238,表明前者结果在“水平”上显著低于后者;另一方面,在20~39岁间各P/F比的全距(即最大值与最小值之差)上,却是前者(0.3780)大于后者(0.2233)。这种结果说明,虽然采用不同口径汇总的调查前12个月内的生育在年龄别生育率与总和生育率结果上并无明显差异,但在应用P/F比分析时,两者可能已经产生不同的内在导向,因为上述两种“正确搭配”的P/F比结果在平均水平和分布上都存在明显差距。同时,对六普1‰样本的P/F比试算结果也产生了新的问题,即第3列和第4列这两种“错误”搭配的结果在表面上均表现得“正常”,即均未出现15~19岁组P/F比呈负值的情况,尤其是“六普口径—普查年龄b”正好对应陈卫、杨胜慧(2014)的计算搭配,但并未呈现“平稳”的结果,而是类似于其他试算结果,P/F比值随年龄明显提高。为什么会出现这样的结果呢?

笔者在比较六普公布的生育率与六普 1‰ 样本按普查年龄对出生分组的生育率值时发现,六普公布的 15~19 岁组生育率要比 1‰ 样本的相应统计相对低得较多,并且该组生育率差异还会影响所有年龄组的累计生育率值(见表 3),可能会构成一个关键影响。因此,笔者根据这个特征简单地将六普公布的 15~19 岁组生育率(0.0059)手工改得稍大一点(比如接近 1‰ 样本相应值的 0.0100),

于是在这样一个微小变化的条件下,即使错误搭配 b 组系数,得到的 P/F 比结果也不再“平稳”,而是比较接近于表 1 中采用 a 组系数所计算的六普 P/F 比值。这一试算结果如表 2 中“六普 b 改”所示。由此可见,P/F 比方法对低龄组生育率(可能包括 20~24 岁组)的敏感性甚至远远大于两套系数的选择,这在应用 P/F 比方法时尤其应当注意。具体到六普公布的低龄生育率值到底有没有质量问题,有哪种质量问题,以及六普 1‰ 样本的低龄生育率相对高于六普公布统计是由于再抽样导致的随机波动或是另有原因,尚无从知晓。

六、试算其他低生育率人口的 P/F 比

P/F 比方法创建于 20 世纪 60 年代,主要是为了解决欠发达国家生育数据少或数据质量差的问题。创建这个方法时曾经用其他众多国家可靠的人口数据来求出将累计生育率折算为对应该次估计(即 P/F 比中的那个 F 值)的两组系数,即前文所说的分别用于按母亲在调

表 2 对六普 1‰ 样本生育率试算的 P/F 比值

年龄组 (岁)	六普口径		五普口径		六普 b 改
	普查年龄 a	普查年龄 b	生育年龄 a	生育年龄 b	
15~19	0.9022	1.7906	0.7966	1.4518	1.4540
20~24	1.0172	1.1855	1.0168	1.1828	1.2791
25~29	1.2577	1.3382	1.2549	1.3357	1.3648
30~34	1.3407	1.3718	1.3397	1.3705	1.3673
35~39	1.3952	1.4072	1.3941	1.4061	1.3873
40~44	1.4700	1.4770	1.4701	1.4771	1.4644
45~49	1.5552	1.5571	1.5543	1.5562	1.5378
20~39 岁均值	1.2527	1.3257	1.2514	1.3238	1.3500
20~39 岁全距	0.3780	0.2217	0.3773	0.2233	0.1082

注:(1)标题中“普查年龄”和“生育年龄”分别代表生育率分子的计算口径,而“a”和“b”则分别代表 P/F 比计算时的系数分组。(2)标题为“六普 b 改”的一列是略微改动了六普公布的 15~19 岁组生育率值后应用 b 组系数的试算结果。

表 3 六普 1‰ 样本按普查时点母亲年龄的生育率及累计生育率

年龄组 (岁)	六普公布统计			六普 1‰ 样本统计		
	平均子女数	生育率	累计生育率	平均子女数	生育率	累计生育率
15~19	0.0125	0.0059	0.0295	0.0122	0.0104	0.0520
20~24	0.2596	0.0695	0.3770	0.2590	0.0736	0.4200
25~29	0.8440	0.0841	0.7975	0.8457	0.0807	0.8236
30~34	1.2928	0.0458	1.0265	1.2918	0.0429	1.0379
35~39	1.5235	0.0187	1.1200	1.5299	0.0184	1.1296
40~44	1.6871	0.0075	1.1575	1.6867	0.0074	1.1668
45~49	1.8366	0.0047	1.1810	1.8396	0.0044	1.1888

查时年龄口径生育率的 a 组系数和用于按母亲在生育时年龄口径生育率的 b 组系数。但是,那时全世界几乎还没有真正意义上的低生育率人口,虽然极个别国家的总和生育率刚刚出现低于 2.1,但低得并不多。因此,该方法创建时确定这两套系数用的人口数据只能取自大多数非低生育率国家。根据统计原理,对这些案例用常规回归求解的系数一般更适用于内插估计,即适用于非低生育率国家。而且采用求解出的系数进行这样的内插估计必定会存在一定的估计误差。因此,数据质量较好的国家不需要使用这种估计方法。这种方法主要用于那些欠发达国家或地区来取得大致的生育水平或调整不准确的生育率统计,并且在那些国家或地区社会发展水平很低,人口的变化也往往很慢,可以近似视为稳定人口。

然而,中国以往几十年经历了迅速的人口转变,在 20 世纪 90 年代初进入低生育水平,并且生育水平和生育模式仍在继续变化。再者,中国拥有海量的人口数据,只是在进入低生育水平时期后才出现对生育数据的强烈质疑,同时也产生了大量统计调整的努力。笔者并不反对用 P/F 比方法对生育水平加以探测,但是,一方面中国人口在变化之中因而并不满足 P/F 比方法的应用条件,另一方面这种方法是否普适于低生育人口还缺乏必要的验证。所以,简单地将 P/F 比方法应用于当代中国,再加上又不了解六普生育率的统计口径,很可能产生误导性极大的结论。

本文尝试将 P/F 比方法应用于其他一些低生育水平人口(见表 4),笔者在计算时全部采用 b 组系数,因为人类生育数据库提供的生育率表的分子明确标志为生育风险暴露年。表 4 中不仅提供了 20~39 岁的 P/F 比平均值,还提供了各个国家或地区生育统计当年的总和生育率(TFR)和 15 年前的总和生育率,并且提供了二者之间的比值和差值。另外,还提供了当年总和生育率与生育率更替水平之间的差值,为了计算方便,将更替水平值一律设为 2.1。

瑞典是以前人口统计学中常用的稳定人口范例,1955~1978 年间瑞典总和生育率从 2.30 下降到谷底(1.60),之后开始回升,1990 年回归至更替水平(2.14),而后出现了第二次下降,1999 年创下最低纪录(1.51)后再次出现回升,2010 年达到 1.99 的新高点后,2011 年又略有下降(1.90)。瑞典这样一个稳定人口范例在低生育水平的进程中如此一波三折,反映了低生育水平进程的复杂性。当代瑞典恐怕很难再被视为稳定人口。瑞典在 1970 年和 1980 年正处于第一波生育水平大幅下降之中,而相应年份的 20~39 岁 P/F 比平均值(1.115 和 1.183)均显著高于 1.0;而 1990 年时刚经历了第一波回升,因此相应 P/F 比平均值略低于 1.0;而 2010 年时则是先经历 10 年的第二波下降然后是 10 年的第二波回升,相应 P/F 比平均值为 1.014。然而,4 个年份的具体 P/F 比值都不太平稳,所以其总和生育率的变化不能视为有生育的漏报或重报。实际上根本不需要看瑞典的 P/F 比值就有理由相信其总和生育率是基本准确的。但笔者是想看看瑞典在生育率变化中年龄别 P/F 比值是否会碰巧出现偶然性的平稳现象,结果是没有碰到。

澳大利亚在 2001 年总和生育率也曾降到 1.75, 在 2008 年回升至 2.0 以上, 之后又略有下降, 2011 年为 1.9。相对而言, 澳大利亚的生育率变化并不太大, 但表 4 中其 3 个年份的 20~39 岁 P/F 比平均值处于 1.138~1.263, 年龄别 P/F 比值也存在明显的波动。表 4 中奥地利和匈牙利均属于欧洲生育率最低的国家。奥地利两个年份的 20~39 岁 P/F 比平均值都较高, 值得

表 4 试算若干低生育人口的 P/F 比值

	瑞典				澳大利亚		
	1970	1980	1990	2010	1986	1996	2001
年龄组(岁)							
15~19	0.9837	1.2140	1.1587	1.0773	1.2666	1.2928	1.3398
20~24	1.1484	1.1750	0.9530	1.1470	1.1504	1.1507	1.2763
25~29	1.0884	1.1699	0.9013	0.9665	1.1001	1.0573	1.2824
30~34	1.1152	1.1925	0.9359	0.9579	1.1377	1.1229	1.1777
35~39	1.1063	1.1947	0.9290	0.9848	1.1658	1.1556	1.3168
40~44	1.0506	1.2366	0.9250	0.9986	1.2317	1.1760	1.2843
45~49	0.9963	1.2574	0.9300	1.0111	1.3564	1.2224	1.2948
20~39(均值)	1.1150	1.1830	0.9300	1.0140	1.1380	1.1220	1.2630
15年前TFR	2.2700	2.4130	1.7690	1.6000	2.9000	1.9390	1.8710
当年TFR	1.9190	1.6800	2.1400	1.9000	1.8710	1.8070	1.7470
比值(前/后)	1.1800	1.4400	0.8300	0.8400	1.5500	1.0700	1.0700
差值(前-后)	0.3500	0.7300	-0.3700	-0.3000	1.0300	0.1300	0.1200
差值(后-2.1)	-0.1800	-0.4200	0.0400	-0.2000	-0.2300	-0.2900	-0.3500
	奥地利		匈牙利		日本	台湾地区	
	1991	2001	2000	2010	1970	2010	
年龄组(岁)							
15~19	1.0519	1.1347	0.5903	0.8106	-84.4164	1.3058	
20~24	1.0938	1.2528	0.8423	1.8356	3.9427	1.7191	
25~29	1.1500	1.2535	1.2019	1.8312	1.2482	1.7351	
30~34	1.1674	1.2598	1.4837	1.1239	1.0519	1.7267	
35~39	1.2017	1.2962	1.4993	0.9917	1.0246	1.8708	
40~44	1.2515	1.3211	1.4343	1.0988	1.1195	2.0598	
45~49	1.2897	1.3762	1.5304	1.1786	1.2974	2.3222	
20~39(均值)	1.1530	1.2660	1.2570	1.4460	1.8170	1.7630	
15年前TFR	1.6880	1.4490	1.7740	1.6420	2.3830	1.7680	
当年TFR	1.5070	1.3300	1.2840	1.3220	2.1290	0.8870	
比值(前/后)	1.1200	1.0900	1.3800	1.2400	1.1200	1.9900	
差值(前-后)	0.1800	0.1200	0.4900	0.3200	0.2500	0.8800	
差值(后-2.1)	-0.5900	-0.7700	-0.8200	-0.7800	0.0300	-1.2100	

注:全表 P/F 比均采用 b 组系数计算。

资料来源:澳大利亚生育数据由澳大利亚统计局张广宇先生提供, 其他国家或地区数据均来自人类生育数据库(The Human Fertility Database.<http://www.humanfertility.org>)。

注意的是, 其 2001 年在 20~34 岁之间的年龄别 P/F 比相对比较平衡, 近似于上述笔者想找的那种案例。按照 P/F 比方法的规则, 可以排除 35~39 岁组, 只计算比较平稳的 20~34 岁 P/F 比平均值, 约为 1.25。但是, 能认为 2001 年奥地利总和生育率(1.330)是由于生育漏报而使之偏低了 25%吗? 至于匈牙利, 其生育率更低一些, 20~39 岁 P/F 比平均值也更高

一些,而年龄别 P/F 比变化非常大。

1970 年日本的总和生育率刚好在更替水平,其 15 年前的水平只略微高一点,因此那时日本正站在低生育率“门口”。但其 P/F 比结果十分怪异,一是 15~19 岁 P/F 比值居然为 -84,二是 20~24 岁的 P/F 比值高达 3.9,结果该异常值将整个 20~39 岁 P/F 比平均值拉至 1.8。人类生育数据库中还有中国台湾地区 2010 年的生育数据,但其 P/F 比结果均高得奇怪。并且,如果只看 20~34 岁区间的 3 个年龄别 P/F 比(分别为 1.7191、1.7351 和 1.7267),也算比较平稳,但如果说三者的平均值 1.7 表明中国台湾地区已经极低的总和生育率(0.887)是由于有 70%的漏报则不会有任何人相信。日本 1970 年和中国台湾地区 2010 年的 P/F 比结果均显得不可思议,笔者并不知道这是出于两地统计口径与其他地方不同的问题还是 P/F 比方法本身的问题。

总之,上述若干低生育人口的 P/F 比测算结果可以归纳出一些共性,那就是 P/F 比方法一般不适用于不断变化的低生育人口。首先,由于进入低生育率后的生育变化仍然可能比较大、比较快,而且还可能出现较明显的时期波动,这种情况会导致 P/F 比方法不适用,若再以此进行间接估计就更没有实际价值。

七、小结与讨论

近年来,人口研究特别是生育研究出现了一种不好的倾向,单纯依赖方法和技术对生育率进行调整,却忽略方法和技术的使用条件,从而出现误导性的结论。政府主管部门之所以在十几年来中坚持不变的 1.8 总和生育率口径,实际上与此类生育率间接估计不无关系。目前政府主管部门和统计部门的生育水平口径都已有了很大变化,而一些人口学者还在用间接估计方法证明生育水平和生育模式没有多大变化。为了提请人口学者注意这个问题,本文通过对中国人口普查生育数据统计口径的讨论入手,说明陈卫、杨胜慧(2014)对六普做 P/F 比测算时因误解六普生育率口径而不恰当地错搭了一组公式系数,并且这种偶然的操作方法失误又巧遇六普公布的 15~19 岁生育率特别低,因而产生了一套看上去很合理、很圆满的 P/F 比统计结果,并使他们错误地断言过去 15 年里中国的生育水平没有出现进一步的下降趋势和六普总和生育率存在 40%的漏报影响。然而,本文通过大量实际数据的试算来探讨 P/F 比方法的特性和局限,这些结果则是从方法的一般性角度反映出 P/F 比方法不太适用于低生育率人口,尤其是在用于调整生育率时要尤其谨慎。本文除了探讨生育率口径与 P/F 比计算系数之间的对应性问题外,还用实例揭示出低龄组生育率值对整个 P/F 比计算结果存在十分关键的影响。因此,P/F 比方法无论是对因偏差、误差导致的虚假特低值抑或是真实的特低值都非常敏感。

根据笔者的理解,低龄组生育率只是累计生育率的一侧边缘,而累计生育率的另一侧边缘应该对 P/F 比结果也有某种类似影响,其实它就是总和生育率水平。实际上,在联合国出版的《人口估计的间接技术——手册 10》中 P/F 比方法部分(United Nations, 1983: 34)曾

提到在将累计生育率折算为对应该次估计的所有年龄组计算公式^①中,总和生育率(标示为 $\varphi(7)$)是作为常数出现的。并且评论说,就理论而言,这类计算公式中有常数项存在并不好,只不过当时在实践中还没有出现过问题。笔者猜测,这也许是由于当时生育水平都在更替水平以上,尤其是P/F比方法主要用于生育率更高的人口。然而,当代许多人口的总和生育率已经不仅仅是低于更替水平,而是达到了超低程度,将P/F比方法应用于低生育人口时,它是否会出现问题就特别值得人口学者警惕。

从另外一个理论视角也可以说明P/F比方法不适用于低生育人口,这就是生育进度效应问题。生育进度效应是与生育模式变化紧密联系的,而存在强烈生育进度效应则是低生育人口的普遍现象。进度效应可以较明显地扭曲时期生育指标,即年龄别生育率与总和生育率,造成时期生育指标的总和生育率值会偏离队列的终身或累计生育指标,因为队列指标反映的是纯粹生育数量,其中并不包含时期进度效应(Bongaarts等,1998)。而在生育转变中,这种效应在所有年龄组都同时出现,各年龄上这种微小变化聚集起来后,对降低时期总和生育率的影响是十分显著的。仅以中国为例,笔者曾根据不同来源的全国调查数据估计出,中国在1990~1995年间推迟生育导致总和生育率平均降低0.11,在1996~1999年间平均降低0.23,在1994~2006年间则平均降低了0.21(郭志刚,2010)。而在P/F比方法的理论框架中,根本不涉及生育进度效应问题,只是简单地将时期的年龄别生育指标 $f(x)$ 累计为总和生育率(TFR),再将其折算为对应该次估计的意义只是为了使生育率这种期中指标值转换为期末指标值(F),就直接与队列生育指标(即平均孩次P)对比。简而言之,P/F比的理论框架带有浓厚的人口统计学早期阶段特征,是将总和生育率等同于终身生育率,因此不适用于低生育时代的生育统计和分析。尽管创建者也说明该方法只适用于生育水平和生育模式没有变化的情况,但并未明确说明年龄别P/F比值表现平稳的结果并不一定能反证实际中生育水平和生育模式没有变化。而陈卫、杨胜慧(2014)正是在这里产生了误解,认为判定是否符合条件并不需要在应用该方法之前就应当考察清楚,而是认为P/F比的计算结果足以反过来证明其应用自身是否合理。

最后,笔者再一次重申对生育率研究中的一种不良倾向的公开批评。这种倾向表现为,企图走捷径、绕过对原始生育数据的深入分析,认为只凭少数几个汇总统计数进行间接估计就能得到“更为准确”的生育率。特别是一些间接估计研究甚至连调查数据的性质和特点还没搞清楚,也不仔细推敲和验证所用方法的假定条件和应用范围,只要发现有些统计数对不上就匆忙判定为出生漏报。这样的急就文章,结论难免出错误。比如,有的研究发现2005年全国1%人口抽样调查数据中育龄妇女申报的前12个月的生育数与样本中0岁人数对不上就简单认为这必是漏报。其实,这种差距显然与当前人口流动量巨大、存在大量留守儿童有关。这些留守儿童与母亲不仅不在同一户生活,甚至都不在同一个省。因此,一个

① 计算公式为: $F(i)=\varphi(i-1)+a(i)f(i)+b(i)f(i+1)+c(i)\varphi(7)$ 。

抽样调查根本不可能做到样本中的0岁儿童正好都是样本中的育龄妇女所生,所以这两个数字对不上也是正常的,要是对应得严丝合缝反而令人不可思议。

应用间接估计法多是由于原始数据极度匮乏、不完整,甚至空白,或者是虽然有原始调查数据但外界得不到(无电子数据),或是只有少数统计结果而又可能很不准确,因而“巧妇难为无米之炊”,既无法评估质量,又不可能深入分析。而我们这里则是拥有大量人口数据,但却被认为是假数,不值得花费精力深入分析。多年来“假数真算”说法影响了人口学者对原始数据的认真分析与挖掘,使原始数据的基础性研究缺乏,而各种间接估计变得时髦。然而,间接估计也需要原始数据汇总的统计指标。如果完全否定原始数据,那么间接估计所用的统计指标凭什么值得信赖,而间接估计取得的结论又凭什么更为准确可信?搞调查统计的人应该知道,伪造几个统计数,或想办法硬凑出某种估计结果,或因巧合得到某种估计结果,其实都远比伪造大量原始数据容易得多。近年来,在国内关于生育水平分析的文献中,对原始数据的挖掘分析已经太少,而间接估计则太多,表明这种倾向已导致对生育水平的分析几乎被生育数据质量评估和生育率间接估计所取代,使生育率研究越来越不接地气,离实际越来越远。

参考文献:

1. 陈卫、杨胜慧(2014):《中国2010年总和生育率的再估计》,《人口研究》,第6期。
2. 国务院人口普查办公室、国家统计局人口统计司(1985):《中国1982年人口普查资料》,中国统计出版社。
3. 国务院人口普查办公室、国家统计局人口统计司(1993):《中国1990年人口普查资料》,中国统计出版社。
4. 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(2002):《中国2000年人口普查资料》,中国统计出版社。
5. 国务院人口普查办公室、国家统计局人口和就业统计司(2012):《中国2010年人口普查资料》,中国统计出版社。
6. 郭志刚(2010):《中国的低生育水平及相关人口研究问题》,《学海》,第1期。
7. 郭志刚等(2014):《2010年全国人口普查的生育数据分析》,载于国务院人口普查办公室、国家统计局人口和就业统计司编:《发展中的中国人口——2010年全国人口普查研究课题论文集》上册,中国统计出版社,第30~74页。
8. 查瑞传、季咏华(1984):《中国妇女生育状况分析》,载于中华人民共和国国务院人口普查办公室、国家统计局人口统计司编:《十亿人口的普查——中国1982年人口普查北京国际讨论会论文集》。
9. Bongaarts, John and Griffith Feeney(1998), On the Quantum and Tempo of Fertility. *Population and Development Review*. 24(2): 271-291.
10. Brass, William and Ansley J. Coale(1968), Methods of Analysis and Estimation. In W. Brass and Others. *The Demography of Tropical Africa*. Princeton University Press: 88-139.
11. United Nations(1983), Estimation of Fertility Based on Information about Children Ever Born. In *Manual X: Indirect Techniques for Demographic Estimation*. New York: United Nations Publication.

(责任编辑:朱 犁)