

人口与发展论坛

人口学研究方法: 规范与发展

本刊编辑部

背景

自 20 世纪 80 年代起, 西方现代的人口分析技术和统计分析方法全面引入中国, 对我国人口科学研究的规范与发展起到了巨大的推动作用。20 世纪 80 年代以来, 中国进行了三次人口普查, 两次全国人口抽样调查, 以及多次有关生育、生殖健康、死亡、迁移和老年等各个方面的专项抽样调查, 人口学研究的数据异常丰富。与此同时, 微型计算机日益普及, 统计分析软件变得越来越易学好用。这些大大推动了统计分析方法和人口分析技术在人口研究中的普及和应用, 取得了丰硕的研究成果。根据郝虹生、陈功的统计, 90 年代以来在《人口研究》杂志上发表的文章中, 定量的分析文章占了主流地位。但是, 在统计分析方法得到推广和普及的同时, 统计分析方法的误用和滥用现象也很严重, 研究方法应用错误和不规范屡见不鲜, 一些教科书上会强调的经典错误, 如把相关关系解释成因果关系, 在实证研究中仍然时有发生。21 世纪中国的人口科学要繁荣和发展, 人口研究方法的规范和研究方法的创新与发展至关重要。本期我们邀请王谦、郑真真和郭志刚三位专家就人口学研究中方法的误用极其规范与发展问题进行讨论, 以期能引起学界和有关人士对这一问题的重视和广泛关注。

认识人口与计划生育统计的特点 避免统计方法和数据的误用

王 谦 (国家计生委计财司 副司长)

由于工作关系和职业习惯, 笔者对各种各样有关人口和计划生育的工作报告、分析文章中的统计方法和统计数据比较关注, 经常可以发现统计方法和统计数据被误用的情况。本文通过一些实例, 分析由于不了解人口和计划生育的特点而造成统计方法和数据误用的情况。

1 个案与群体

〔实例 1〕某调查报告称, 某村 2001 年出生婴儿 10 人, 其中男婴 6 人, 女婴 4 人, 出生性别比高达 150, 严重失调。

〔实例 2〕某乡给某村下达人口计划, 其中一项指标是计划生育率, 要求当年的计划生育率不低于 95%, 而该村一般每年出生不足 20 人。

〔实例 3〕某地计生委根据群众举报, 查出某县一个超生 5 胎的情况, 据此称该县超生问题严重。

〔实例 4〕某县总人口不足 50 万人, 近年来在孕妇中推广服用“福施福”。上级要求检查服用“福施福”后, 人口缺陷发生率是否逐年下降。还有不少地方把孕产妇死亡率是否逐年下降当作生殖健康服务工作的考核内容。

这几个例子所出现的问题都是以个案或少量发生的情况, 说明一个地方宏观的情况。人口和计划生育统计所分析的对象(人口)或事例具有群体性, 这个群体是由每个个体的人或事件集合而成。每个个体是否发现这种现象是偶然的, 作为表现群体规律的统计数据, 只有当群体具有相当的规模后才有意义。群体的规模太小, 尽管对每个个体的统计是准确的, 指标的计算也是正确的, 但计算结果并不能客观地反映群体的规律。

如实例 1, 一个村的出生人数仅为 10 人, 计算的出生性别比要么是 100(“绝对平衡”), 要么就是男女比例严重失调。实例 2 中计划生育率受出生总数的影响, 要么计生率正好 100%, 只要有一个计划外出生, 计划生育率就低于 95%。例 4 涉及的出生缺陷发生率或孕产妇死亡率一般统计时均以 10 万作为分母, 对于一个县, 每年出生几千人或 1 万多人, 即使按照出生缺陷、孕产妇死亡发生的平均水平, 每年也只有几例, 偶然性很大, 在统计数据上很难表现为逐年下降。至于例 3, 以一个特例说明全县的情况, 更是欠妥。各种人口和计划生育统计教材上并未说明各项统计指标适用的人群至少要达到多少, 但一般说来这些统计指标绝大部分不能用于分析规模在几百人到一千多人的村级情况, 许多指标在县、乡级使用也不太合适。多年来各级是通过报表收集人口和计划生育统计数据。报表由上级制发, 基层单位按统一的口径填报每个项目, 然后逐级汇总、上报。这样以来, 上、下级的报表式样相同, 只是数量大小有差别, 于是, 上级单位计算哪些指标, 下级单位“照葫芦画瓢”计算相同指标。特别是实行人口与计划生育目标管理责任制后, 各项工作任务和责任要逐级分解、落实, 相应地, 工作指标也被层层分解, 上一级控制的指标(如出生率、计划生育率)也被一直套用到基层单位。在这种情况下, 群体性的要求很容易被忽视。

2 自然属性与社会属性

[实例 5]1995 年 8 月, 在全国上半年人口形势分析会上, 某省计生委分析本省当年上半年二孩出生数量比上一年同期减少的原因时称, 由于自当年起全省广泛推行“三结合”, 许多群众为了发家致富, 主动退出二孩指标, 于是二孩出生明显减少。笔者的同事立即对此理由提出了置疑: 既然是主动退出二孩指标导致二孩出生减少, 那么原来应在 1995 年上半年生育二孩的妇女至少在 1994 年上半年就应退出二孩指标, 而那时候, 该省并未开展计划生育“三结合”; 至于 1995 年上半年开展“三结合”后退出的二孩指标, 其二孩出生减少的效果应在 1996 年以后才能表现出来。群众退出二孩指标与二孩出生数量的减少在时间上出现了“矛盾”。

[实例 6]许多文章分析妇女受教育程度越高, 生育的孩子越少。

这两个例子说明分析人口出生数量的变化应考虑人口的自然属性和社会属性。人口首先是生物意义上的人口, 具有自然属性, 同时, 人口生活、存在于一定的社会环境中, 具有社会属性。人口现象的发生既受自然属性的影响, 也受社会属性的影响。从根本上说, 社会属性对人口现象的影响和制约是有条件的、间接的。而自然属性的影响和制约是无条件的、直接的。社会属性的影响和制约常常要通过自然属性来实现。在分析人口现象的时候, 首先要解释人口的自然属性, 然后再解释社会属性。就实例 5 而言, 从妇女退出二孩指标到二孩出生减少, 至少间隔 10 个月, 即妇女的怀孕期, 无论推行“三结合”力度多大, 妇女 10 个月的怀孕期不能缩短。退指标的发生与出生数的减少在时间上不“同步”, 至少要“滞后”一年。也就是说, 1995 年上半年因开展“三结合”妇女退出二孩生育指标不会影响这个时期的二孩出生数量。关于实例 6, 一个人的受教育程度属于社会属性, 妇女受教育水平提高, 的确会影响其生育行为。然而, 影响生育的自然因素, 直接因素是避孕节育行为, 具体地说, 影响生育的生物因素包括推迟结婚(不发生性行为)、采取避孕措施和流产(终止妊娠)。妇女文化程度影响到上述生物因素发生变化(如文化程度高, 能够更好地掌握避孕知识, 可以提高避孕的有效性, 减少怀孕的可能), 并通过上述因素影响生育孩子的数量。直接讲文化程度与孩子数量的关系, 未免有些牵强。

3 定量与定性

[实例 7]各地每年都要分析当年的人口形势, 将当年的数据与上一年的数据进行对比。许多分析报告在列举了对比数据后得出结论: “今年的工作比上一年有明显的进步, 取得了巨大的成绩, 上了一个新台阶”。但如果我们仔细察看各项指标的数量, 发现变化并不大, 如计划生育率由 88.5% 上升到 89%, 出生率由 15.4‰ 下降为 14.9‰(这其中由于人口年龄结构的变化也能导致出生率下降), 多孩出生由 165 人减少为 149 人, 多孩率由 2.2% 下降为 2% 等等。根据这些数据的变化, 得出的定性结论应

当是:“工作稳步发展,人口增长保持平稳的态势”。人口和计划生育的统计分析主要是指对数据的分析,并由“定量”的描述引申到定性的判断。如果忽视定性的分析,那么,所谓定量分析不过是数量变化的文字描述而已。把握好定性分析的关键在于,定性分析应以定量分析为基础,即“定量”在前,“定性”在后。那种先“定性”,再计算数量变化的做法,违背了“定量”与“定性”的基本规律。这样进行的定量分析,只是给定性的结论加一些数量的点缀。

4 模糊和精确

〔实例 8〕某地区的一份材料称,根据 20 年来每年总和生育率的变化,推算本地区 20 年来少生了 1831275 人。

〔实例 9〕某县计生委利用农村赶集日人口比较集中的机会,出动服务车,设置咨询台,向群众宣传计划生育知识,提供咨询服务。县计生委的汇报材料称,“一个月来在集日上接受宣传、咨询的群众累计达到 127328 人次”。

乍一看,实例 8 和实例 9 中的人数和人次数统计和计算十分认真仔细,数字如此精确,令人钦佩。但转念一想,不免生疑:实例 8 中利用总和生育率计算少生人口,需要以年龄别生育率与分年龄妇女人数相乘计算出生人数,一般来看,每个年龄组的计算结果都不会是整数,只好四舍五入。每一年若干个年龄组累计下来,个位数已经含糊不清了。再把 20 年的少生人数累计起来,怎么可能精确到个位数呢?况且 20 年间,该地区的人口并非处于“封闭状态”,各年龄组人数因为人口迁移而变化,也会影响到计算少生的数量。至于实例 9,在一个开放的、人们频繁走动的集市上,如何判定哪个人接受了宣传,哪个人接受了咨询,并没有一个客观的标准;即使有标准,实际操作中也很难把握,因此,根本没有可能逐人统计。实例 8、实例 9 中看似精确的数据令人感到滑稽,倒不如模糊一些,说“20 年大约少生了 18 万人”,“一个月在集市上接受了计划生育宣传、咨询服务的人次达 10 余万”,或许有几分可信。从某种意义上讲,人口和计划生育所涉及的是宏观层面的事情,与此相应的统计数据所反映的是大致的状态或趋势,没有必要也没有可能做到精确。许多同志在应用统计方法时,往往注意的是方法本身的定义和计算过程,而对计算中数据的来源不了解。人口和计划生育统计属于社会经济统计范畴,很多数据是通过抽样调查获得,允许有一定的误差;即使是全面的调查(如人口普查、全局报表),在实施过程中,也会遇到这样、那样的干扰或影响,存在着调查误差,这是无法完全避免的。从某种意义上讲,数据存在一定的误差,这是人口和计划生育统计的特点之一。

5 原因与结果

〔实例 10〕最典型的例子莫过于“人口出生率每降低 1 个百分点,人均 GDP 提高多少个百分点”的说法。这种说法的依据是利用相关分析法对全国各省的人口出生率和人均 GDP 的数量变化进行计算,得出二者之间的函数关系式和相关系数。把人口出生率作为自变量,看自变量的单位变化影响因变量(人均 GDP)变化的数量大小。

在这个例子中相关分析法本身以及有关的数据都是正确的,然而,用这样一个函数式来解释人口增长与经济成长的相互关系是有问题的。其一,就统计方法来说,相关分析只反映变量之间的相关关系,而不说明因果关系;其二,人口增长与经济成长固然相互影响,但从本质上来说,是经济发展影响人们的思想,进而影响人们的生育行为,最终导致生育率下降,人口增长率降低,即经济成长是“因”,而人口增长是“果”。颠倒了因果关系,不仅逻辑上讲不通,数量上的“相关性”也失去了意义。

笔者在多年的工作中体会到,由于人口和计划生育所具有的特点,人口和计划生育统计分析方法的应用和统计数据的解释有别于其它部门和领域对于统计的一般要求。正确地应用统计分析有助于我们认识人口和计划生育的特点和规律,只有科学地认识和把握人口和计划生育的特点和规律,才能避免在统计分析中陷入误区。

社会科学实证研究中的统计分析方法应用

郑真真 (北京大学人口研究所 副教授)

统计学的应用随着微型计算机的普及越来越广泛,在社会科学实证研究中几乎是无处不在。有了一定规模的数据和一个统计分析软件,就可以很方便地进行各种估算和分析。然而由于统计分析方法本身并不像加减乘除那样简单,而一些统计分析软件已经发展到几乎是人人都可使用的程度,如果使用者在只知其然不知其所以然的情况下操作并得到结果,可能出现对统计分析方法误用或滥用的现象。本文仅对一些统计分析中比较常见的问题进行讨论,以引起各方面的重视。

1 描述性统计

描述性统计是社会科学实证研究中最常用的方法。准确、全面、正确的描述是所有实证分析的基础,如果对某个事件或某种现象的描述不清楚或存在偏差,那么其后的所有分析都将是值得怀疑的。一项研究能够将所研究的现象或对象描述清楚,就是一个极大的贡献;而描述的偏差可能会引起公众或学术界对某些社会现象的误解,甚至误导政府决策。但是因为描述性统计所用方法简单易得,往往没有得到足够的重视。

均值的局限 普遍用于描述样本集中趋势的测量之一是均值。它对于近似正态的对称分布样本来说是比较好的测量,对于不对称分布则不然,尤其会受到极端值的影响。两个分布完全不同的样本可能会有相同的均值,因此均值在某种程度上抹杀了样本内部的差异,而往往这种内部差异正是需要我们进行深度研究的、或应当引起人们注意的。为了弥补均值的这个缺陷,一般在报告均值的同时也报告方差,或用直方图/散点图的形式描述分布,以提请读者注意群体内部的差异。

不同群体的可比性 在描述性统计中,往往涉及到对不同时期或不同人群的总体描述,以反映社会变化或地区差异。在社会科学中、尤其是人口研究中,不少事件的发生都是与年龄密切相关的,如我国妇女大部分在35岁以前完成了生育,从而导致35岁以上育龄妇女中极高的避孕现用率。在这种情况下,两个样本之间存在避孕现用率的差异可能只是年龄结构的差异,而不是年龄别避孕现用率的差异。又如在报告流动人口犯罪问题时,给人的印象往往是流动人口犯罪率高于常住人口,但忽视了流动人口的年龄和性别构成与常住人口完全不同,且青年男性是犯罪率较高的人群。这种对两个不同群体的比较往往会导致错误的结论。

绝对数的使用 由于中国人口数量巨大,调查研究也比较容易得到大容量的样本,所以对任何小概率事件用绝对数报告都会出现惊人的巨大数字,单纯对绝对数的强调往往会产生戏剧性的效果。比较合理的方式一般是在报告某事件绝对数的同时,给出该事件的发生率或占研究人群的比例。

小样本的代表性 在一次抽样的小样本中求得的率或比例会非常不稳定,与另一次抽样的结果可能会有较大差距。因此当研究仅限于从小样本获得的资料时,应当在报告比例的同时也报告样本量。

2 双变量统计分析

在社会科学研究中,首先分析的往往是两个变量之间的关系,如用相关或列联表等方法。一般在确定两个变量之间确实有某种关系,如在经过统计检验后证实两变量有显著相关关系,进行更进一步的分析才有意义。因此,双变量统计分析在实证分析中占有重要地位。但是,由于在应用中对有些问题的忽视,双变量统计分析也很容易出现偏差或错误。

卡方检验的局限 在利用列联表对两个定序/定类变量进行相关分析时,需要进行统计检验来判断两个变量的相关是否有统计上的显著意义。不少研究结果都用卡方检验的显著性报告相关状况。但值得注意的是,卡方统计量的计算本身是有局限性的,样本越大,卡方值就会相应增大,因此大样本的卡方检验很容易得到显著结果。所以一般在报告卡方检验结果以说明两变量是否显著相关时,还

应当同时报告相关强度, 即相应的相关系数, 如 Γ , Λ 等。

统计意义上的显著与差别的实际意义 在检验两个定距变量的均值差别是否具有统计上的显著性时, 也存在相似的问题。由于样本量越大, 样本均值分布的方差就越小, 因此常用的 t 检验结果就越可能显著, 任何细微的差别都可能具有统计上的显著性。但有时具有统计意义显著性的差异, 在实际生活中可能意义并不大, 如同在两个草堆之间找出一根草的差距, 对判断两个草堆的大小没有实际意义。因此, 对任何检验结果都应当有符合实际的解释和说明。

虚假相关问题 双变量分析中的虚假相关问题, 几乎在所有关于社会科学研究方法的教科书中都会涉及到, 在统计分析方法的教学中也被视为经典问题。但是多少年来, 人们仍然在不断地重复着这个“经典的错误”, 即认为可见的或统计检验结果显著的相关就是真正的相关; 更为大胆的做法是把这种相关关系推向因果关系。我们知道, 对于有的变量来说, 即使是经过检验判定两者具有统计上显著的相关关系, 也不一定存在实际意义上的关系, 因为可能有未考虑到的变量或不可测量的变量在同时对两个研究变量起作用, 有时甚至可能完全是偶然的巧合。例如, 火灾的大小是以火灾损失来衡量的, 而参加灭火的消防员人数是与火灾大小有关的, 火灾越大, 出动的消防员就越多, 但凡是具有常识的人都不会根据出动消防员人数和火灾损失两个变量之间的高度相关, 断定出动消防员越多火灾损失就越大, 因为火灾的规模是决定因素(但很难直接衡量)。在有关人口科学研究中也有报告虚假相关的现象, 如人口增长率的降低导致了经济增长的提法就是一例。因此, 在分析相关关系时, 应当根据理论、知识、经验、甚至常识来判断这种分析是否有意义、是否存在其他变量的作用(称为外在变量), 避免得出有悖于常理的分析结果。有些虚假相关是可以通过统计分析方法判别的, 如在控制了另外一些变量后观察两个变量的偏相关, 或在双变量分析的基础上, 进一步用多变量分析深入研究。

3 多变量分析

回归分析是多变量分析中应用最多的方法, 尤其是逻辑斯蒂回归更是被广泛地应用。在众多应用中, 比较明显的问题是使用方法是否得当和对结果的报告和解释是否规范、合理(见 2002 年第 2 期《人口研究》刘金塘文)。此外还有一些应当引起注意的问题。

分析框架的重要性 在社会科学研究中, 各变量之间往往存在错综复杂的关系, 如果在进行回归分析之前没有一个清晰合理的分析框架, 那么回归的结果有可能会引起质疑。一般应在报告回归分析结果之前, 介绍该分析的框架, 如各变量的定义、各自变量与因变量的假设关系及其理由等, 对建立的回归模型做出合理性论证。有一些变量可能是作为控制变量纳入回归模型的, 如性别、年龄等, 最好事先解释清楚。对假设因果关系的模型, 应当至少能够说明: (1) 该因果关系在理论上是正确的、在实践中是合理的; (2) 从事件发生的时间上来说, 应当是原因发生在先, 结果发生在后。如有些回归分析中, 未加说明即把所有与因变量显著相关的变量都囊括在自变量中, 甚至有些自变量与因变量有明显的互为因果关系, 显得分析逻辑混乱; 还有的论文在简单介绍研究背景和数据来源之后, 急于建立因果关系并推出回归分析结果, 然后再根据各变量在回归模型中的显著性一一说明, 这相当于事后解释; 这些做法都是错误的。

在具备“奔 4”微机和较易操作的软件的今天, 转瞬间就可完成一次回归分析, 但是在此之前, 需要有大量的前期准备工作, 包括文献检索和理论框架构建, 才能确保统计分析的科学性。

分析方法应用的条件 每种多变量方法都有各自的前提条件或假设, 如果这些条件不具备或者假设不成立, 该方法的应用就成问题。如 Pearson 相关是考察线性相关关系, 多元方差分析只能辨别线性相关因变量的多元差异, 线性回归分析假设自变量与因变量之间为线性关系, 因子分析方法也是建立在各变量具有一定的线性相关基础之上的; 另外, 在逻辑斯蒂回归中, 每个分类都应保证有足够的频数, 如果频数太少就会影响参数估计的稳定性; 等等。尽管一般不在报告分析结果时说明各种假设是否成立或条件是否满足, 但是在进行分析时应当自觉地进行考察。如果不能满足条件或假设不能成立, 就对数据进行转换或调整后再分析, 或者改变分析方法。

多变量分析结果的展示和解释 多变量分析的结果一般是通过列表来展示的。现在一种并不少见的做法是直接把统计软件的输出直接复制到论文中,我们往往会在文章中看到包括回归参数估计、参数标准差、检验统计值、检验显著性、偏相关系数等等 n 行 m 列的大表,使人有目不暇接的感觉。实际上参数标准差和检验统计值是提供给分析者的信息,没有必要列在结果中;如果不是有特别需要的话,偏相关系数也不是关注重点;最主要的应当是回归参数估计及其显著性。

在列出分析结果之后,应当对结果的实际意义进行解释和讨论,而不是复述分析结果的数学意义。此外,在多元统计分析中一个常见的问题是分析者对变量作用不具有预期统计显著性的失望,因此绕开不显著的变量,甚至对数据或模型进行各种调整以获得显著结果。其实,统计分析结果不显著往往也是有实际意义的。例如在分析我国高龄老人的地区分布时发现,高龄老人比例与当地医疗卫生指标没有显著关系,这说明我国医疗系统还没有具备延长老人寿命的功能;另一方面也说明这些高龄老人的存活不是主要靠医药维持的。所以,在解释分析结果时,只要是在分析框架中涉及并参与分析的变量,无论作用显著与否,都应当给予充分的讨论;对于那些由于知识或信息的限制难以得出结论的结果,可以作为问题提出,以便进行更有针对性的进一步研究。

此外,任何方法都有其局限性,分析结果也不会十分完美。因此在讨论结果的同时,也应当就此向读者说明。例如当一个多元线性回归分析的确定系数较低时,需要指出该模型有限的解释能力,探讨可能存在但没有纳入分析的更重要的影响因素。

不必求最新、只求最合适 有些研究生在撰写学位论文时,常常因为自己没有应用最新的统计分析方法而感到忐忑不安;在评论某项研究的创新性时,有时也出现把学术创新和应用新方法混为一谈的现象,例如认为应用描述性统计方法的研究水平低于应用解释性或预测性方法的研究。新方法是层出不穷的。但是,出现了新方法并不意味着传统方法就不再适用,而是各有千秋。统计分析方法是工具,哪件合适就用哪件,能用锤子解决的问题不必开冲床。有时越是复杂的方法,假设条件也会相应较多,应用的局限性更大。因此,盲目追求方法的新颖并不是高水平研究的保证,真正需要注意的是使用最合适的方法。而对所用方法的真正了解,是正确运用统计分析方法的前提。

总和生育率的内在缺陷及其改进

郭志刚 (北京大学社会学系、北京大学中国社会与发展研究中心 教授)

1 总和生育率的应用目的及评价原则

年龄别生育率及其概括性指标总和生育率(TFR)是生育研究中最常用的指标体系。但是现在它们面临很多实际问题。本文不讨论出生漏报导致统计失实的问题,因为其性质并不在于统计方法,而是一个社会问题。本文只局限于这一指标体系内在的有效性问题的方法论讨论。一个指标是否有效应该以研究目的为标准来判断,因此这一讨论将结合当前实际工作的需要进行。

统计指标有两种功能:一种是对调查对象本身特征的直接描述,另一种则用于推断估计。很多情况下,统计指标同时承担着这两种功能,比如样本统计量反映了样本对象的特征,同时又服务于推断估计总体参数。既然是一种估计,自然会有推断偏差或误差。评价不同估计的优劣是看谁的偏差或误差更小。

那么通常我们应用 TFR 到底要反映什么?一是为了在控制育龄妇女结构的条件下概括时期生育水平,二是作为终身生育水平的估计。^①两种性质都反映在各年龄组生育率的合计上。通常,不同基

^① 在这一方面,总和生育率与人口的粗再生产率与净再生产率的性质类似,后两个指标中只不过又控制了下一代的性别和死亡因素的影响而已。

数的相对数指标不可以直接相加^①。为什么年龄别生育率可加,是因为设置了假设队列的概念。TFR 主要不是平均的概念^②,而是总和的概念(即假设队列经过所有年龄后生育总和)。因此,TFR 一身兼两任,其内容为时期生育率总和,其形式为队列终身生育水平。从前一种意义上使用时描述了该时期生育水平,从后一种意义上使用时则是假设队列终身生育水平的估计^③。其值实际上涉及了 35 个实际队列,并且只有在生育水平及年龄模式长期不变的苛刻条件下,它才真正与实际队列终身生育水平相吻合(但是队列仍是泛指)。尽管这一假设队列与真实队列并不能很好对应,但起码可以及时提供一种队列终身生育的估计来满足实际需要。

2 缺陷一: 总和生育率对终身生育水平的背离

表现 A: 最早对 TFR 的批评是由于其剧烈的时期波动。一逢时期突发事件(如中国 1958~1961),实际生育量发生了变化,TFR 就会大幅度下降。应该说,批评并不是指向其描述时期生育水平的功能,而是指向其作为终身生育估计的功能。因为,这时 TFR 的下降只是由于时期特殊原因影响,并不意味着终身生育水平真的下降。实际上,时期效应一过,TFR 马上便会出现反弹(即常说的生育补偿),然而反弹的水平也并不能标志终身生育水平真的那样高。总之,TFR 短期内剧烈波动时,将其作为终身生育率估计来理解很成问题。其中最引人注目的是,反弹年份的分孩次 TFR(i)会超过 1,特别是一孩 TFR(1)会大大超过 1^④。人口统计学对此采取的对策是,避免采用这些年份的 TFR 来作为终身生育水平的估计,或者采用若干年份的 TFR 的平均值来作为终身生育水平的估计,希望将欠年与盈年的误差相抵消。

表现 B: 然而,有时即使社会中似乎并没有什么特殊事件,TFR 也会背离终身生育水平。但是,往往只是在 TFR 提高时才会受到一定关注^⑤。理论分析可以证明,这种背离既可以是正的,也可以是负的。特别是在负偏离的情况下,往往可能持续较长的时间。并且,这种负背离实际上正是当前所面临的实际情况,因此特别需要重视。

这种背离产生的原因是婚育年龄的变化,或者说是队列的年龄别生育模式的改变。而负偏离则对应着婚育年龄的推迟。与时期突发事件对婚育年龄的推迟的爆发性影响不同,生育模式的自身转变具有较长时期的持续性、变化上的渐进性、现象的隐蔽性等特征,并且它并不一定伴随生育补偿现象。甚至有时人们根本没有意识到这种偏离的存在。

人口统计学早就揭示出,晚婚晚育可以延缓人口增长,但这是从长期人口发展的角度来证明的。而晚婚晚育对年份 TFR 有什么影响,则很少有人涉及。从理论上说,取得同样的终身生育总量,可以有不同的进程表。即假定各队列的终身生育数量不变,从某一时期开始各队列的生育年龄开始推延,这一过程可以持续较长一段时期,直至最终稳定在一个新的生育模式上。我对此做过一些模拟计算,在这种情况下,生育模式转变时期中每年的 TFR 一定会低于事先所设的队列终身生育量,而这段时期两端及以外各年份的 TFR 则可以等于事先所设的终身生育水平。这表明,即使将整个转变过程或更长时期中各年的 TFR 都加在一起平均,得到的还是一个偏低的终身生育估计。

更普遍的情况是生育模式的变化与终身生育量的变化同时发生,这时 TFR 下降同时受这两个因素变化的影响,由生育模式变化导致 TFR 对终身生育水平的偏离便较难分析。

① 如男性平均年龄加女性平均年龄没有意义。

② 早期将其翻译为平均生育率的确是错误的。至于年龄别生育率对妇女其他差别所做的均质假设,其实存在于所有汇总指标之中,而不论其分组有多细,但是都不会因此而称为“平均××率”。

③ 但这种估计从原理上不同于一般统计量从样本推断总体的情况,但有关评价原则却是类似的。

④ 1982 年全国 1‰人口生育率抽样调查数据显示(姚新武,1995),1963 年 TFR=7.463,其中各孩次 TFR 都大于 1,最高的是 TFR(1)=1.568。作为队列估计,即是说每人生育一个半一孩,显得很荒唐。

⑤ 如 1983 年左右中国人口学界对 TFR 的讨论。

实际问题: 实际工作需要不允许我们等很多年再提供真实队列终身生育统计, 而现在只有 TFR 一种估计方法, 并且我们知道 TFR 还会经常偏离队列终身生育水平, 因此急需寻找一种更好的估计来取代它。否则尽管当前 TFR 的值虽然很低, 即使不论统计失真问题, 我们也不知道这是否仅仅反映 TFR 发生了负偏离, 那么我们凭什么肯定生育率已经下降到更替水平了。我们又怎么能知道, 当生育年龄推迟告一段落时, TFR 向终身生育水平回归时会回升多少。要知道更替水平的真正概念是实际上某队列与其终身生育后代数量的比^①。

有关改进: Bongaarts 和 Feeney (1998) 提出了去进度效应总和生育率 (TFR')。该方法旨在提供一种根据时期数据对终身生育水平的较好估计^②。郭志刚 (2001) 对此进行了介绍, 并用中国多年生育数据对该方法进行了检测和评价, 结果是 TFR' 作为终身生育水平的估计的确大大优于 TFR^③。

该方法基本原理可以这样来理解: 现实中生育模式转变时, 除了导致该年生育数量有所变化以外, 还会有其他共生现象, 如分孩次的平均生育年龄 (MAC_i) 也会变化^④。MAC_i 实际上是生育模式转变的测量值, 可以在理论上建立其变化量与时期生育变化量之间的函数关系。TFR' 便是在常规分孩次 TFR(i) 的基础上利用 MAC_i 的变化信息来调整, 得到去进度效应的分孩次 TFR' (i), 然后再汇总为 TFR'。经过调整, TFR' 可以在相当程度上修正 TFR 距终身生育水平的偏离, 因此 TFR' 更接近于队列终身生育水平 (这里队列仍是泛指)。也就是说, 我们可以用 TFR' 来替代 TFR 原来所承担的终身生育估计的功能, 而 TFR 还可以继续承担描述时期生育水平的功能, TFR' 与 TFR 之差可以作为生育推迟对当前生育水平影响的估计^⑤。

我认为, 尽管 TFR' 指标还有继续改进的余地^⑥, 但显然 Bongaarts 和 Feeney 为解决这一估计问题指出了极富于科学价值的方法论途径。并且, 这一新指标不仅可以用于监测队列终身生育水平, 也可以用于分析以往的生育数据 (郭志刚, 2000), 帮助我们更好地理解我国的生育转变史和计划生育史。

3 缺陷二: 总和生育率不能控制育龄妇女的孩次结构

对某些重要因素进行统计控制可以为不同年份或不同地区的比较提供更大的可比性。尽管 TFR 控制了育龄妇女的年龄结构, 但是它并没有控制育龄妇女的孩次结构。然而, 当前许多人口模拟研究都不能再忽略育龄妇女的孩次结构问题, 否则会有损其研究结果的有效性^⑦。

问题的表现: 实际上生育过程是严格的递进事件, 也就是说, 只有未生育者才能生育一孩, 只有生育过一孩且尚未生育二孩者才能生育二孩。而常规生育率则是以年龄别所有妇女总数作为基数, 因而没有容纳育龄妇女的曾生孩次信息, 也就不能控制曾生孩次这种结构影响。因此, 比较孩次结构差别较大的年份或地区之间的年龄别生育率 (或 TFR), 便不能区分其差异到底是出于生育水平不同还

- ① 从这个意义上, 净人口再生产率等于 1 仅是一种时期估计的标准, 并且也服从上述偏离, 并不是真正的更替水平。
- ② Bongaarts 和 Feeney (1998) 警告说, 这一方法不适用于那种特殊时期效应 (即重大灾害等) 的年份。
- ③ 郭震威 (2000) 认为 TFR' 不伦不类, 既不是时期生育指标, 又不是终身生育指标, 是个尴尬的指标。我认为, 它只是应实际需要而产生的另一种根据时期生育信息对终身生育水平的新估计而已, 其实这没有什么可尴尬的。统计中凡是不能直接测量的时候, 都得采用估计来代替。真正有意义的问题是, 它是否比传统 TFR 更接近于终身生育水平。
- ④ Bongaarts 和 Feeney (1998) 强调要用分孩次的平均生育年龄, 而不能用总的平均生育年龄, 因为后者会抹煞实际变化。
- ⑤ 我曾当面请教 Bongaarts, 问 TFR' 方法是否可应用于预测模拟, 他毫不犹豫地回答说不能。
- ⑥ 比如某一年的 MAC_i 的计算还要前一年和后一年的数据, 因而多少丧失了一点及时性。然而, 我自己曾尝试过另外的计算, 比如只用前一年和当年两年的信息, 调整的结果也并无差。
- ⑦ 比如, 与调整生育政策相联系的人口模拟如果忽略现有一孩的妇女人数, 便无法计算由于多年积累在一孩的这批妇女在政策调整时将产生的特殊生育效应。另外, 有关生育政策后果方面的研究 (如四二一家庭结构) 也有类似问题。

是出于孩次结构不同。当其用于人口预测模拟时,如果假定两地区育龄妇女年龄结构相同,同样的系列生育率的假设会导致孩次结构十分不同的地区(代表着不同生育政策类型)将会有相同的出生人数,显然这严重地偏离实际情况。因此,面对当前的许多研究需要,基于年龄别生育率的 TFR 方法无论是用于比较分析上还是用于预测模拟,都会影响其结果的有效性^①。

替代方法的产生:实际上,计划生育采取了分类指导原则后,便产生了多种生育政策类型,其显著特征便是孩次控制上的不同。在不同地区实施不同政策时,育龄妇女的孩次结构便取得了越来越重要的意义。从这一角度出发,导致 Feeney(1985)和马瀛通等(1986)分别以不同思路创建了孩次递进生育指标体系。

他们的共同点在于采用孩次递进比作为生育测量的基础来构建指标体系。如前所述,每年的出生可以划分孩次。如果再将某一群体生育指标的基数由对应口径的年中妇女总数换成年初对应(即前一)孩次妇女时,便称为孩次递进比^②。然后,还可以将孩次递进比换算成递进总和生育率。

Feeney 的模型突出地考虑了孩次递进间隔因素,不考虑年龄结构,因而常被称为间隔递进模型。而马瀛通等的模型则突出地考虑了年龄递进因素,因此常被称为年龄递进模型。

这两种方法不仅可用于在控制妇女孩次结构条件下的比较分析,也可以用于预测模拟^③。递进方法不仅可以在预测中控制孩次结构,而且应用于模拟预测时实际上比常规生育率方法更方便^④。由于递进比的概率性质,还能根本避免孩次别总和生育率出现的那种超过 1 的现象^⑤。

参考文献:

- 1 刘铮, 邬沧萍, 查瑞传. 人口统计学. 中国人民大学出版社, 1981
- 2 马瀛通, 王彦祖, 杨叔章. 递进人口发展模型的提出与总和递进指标体系的确立. 人口与经济, 1986; 1, 2
- 3 姚新武. 中国生育数据集. 中国人口出版社, 1995
- 4 郭震威. 对“去进度效应总和生育率(TFR')方法”的一点看法. 人口研究, 2000; 1
- 5 郭志刚. 从近年来的时期生育行为看终身生育水平. 人口研究, 2001; 1
- 6 郭志刚. 时期生育水平指标的回顾与分析. 人口与经济, 2000; 1
- 7 Feeney. 1985. Parity progression projection. International Union for the Scientific Study of Population, International Population Conference, Florence 1985, vol. 4
- 8 Bongaarts and Feeney. 1998. On the quantum and tempo of fertility. Population and Development Review 24(2): 271~291/ 中文译文: 人口研究, 2000; 1

主持人评论

本期论坛邀请的三位专家都有多年从事人口统计和数据分析的经验,对人口学研究方法有很深的造诣。王谦副司长通过从多年工作中积累的大量实例,分析了在人口和计划生育工作中经常发生

- ① 至于分孩次的年龄别生育率,只是对分子(即出生)划分了孩次类别,但是并没有对育龄妇女本身划分孩次结构。其基数没有控制孩次结构,因而从本质上说也没有控制孩次结构。
- ② 实际上可理解为一种概率,如本年年初某年龄组(或孩次间隔组)只有一孩的妇女中在本年生育二孩的比例。
- ③ 应该指出,其实中国人民大学人口研究所的老一辈人口学家早在 1970 年代所自行创建的标准(或可变)生育率预测法中也已经包含了控制孩次结构影响的思想(刘铮、邬沧萍、查瑞传, 1981)。
- ④ 这里的方便指,它只需要每年期初妇女的孩次别人口数便可以直接乘以递进比便得到对应的生育量,而不需要再计算其中妇女人数。
- ⑤ 尽管如此, Bongaarts 和 Feeney(1998)认为这一类应用生命表方法的生育指标仍然没有真正控制生育进程变化的影响。

的统计方法和统计数据被误用的情况;郑真真副教授则对统计分析中比较常见的问题进行了独到的分析和总结;郭志刚教授对总和生育率指标进行了深入全面的剖析,分析了其内在缺陷和改进方法。

王谦副司长通过 10 个典型实例说明在人口和计划生育实际工作中,统计方法和统计数据在五个方面经常被误用。这五个方面的问题有些属于统计分析中容易犯的错误,带有普遍性,如,错误地解释变量之间的因果关系,在统计分析中忽视定性分析与定量分析的关系,结论不是由统计分析中得出,使得定量分析只是定性结果的一种点缀。有些问题则由于人口与计划生育领域的特点而显得更为严重,如他指出的,大部分人口和计划生育统计指标在村一级或县、乡不适用是因为总体规模不够。统计所研究的是总体的数量关系而不是个别的数量关系,所以统计指标在一定的总体规模下才有意义,这是统计学中的基本概念,但是在实际中却经常被误用。我们经常看到的错误都是统计学上的一些“基本概念错误”,而不是“高深错误”。统计学上的概念和方法可能很简单,也容易理解,但是要在实际工作中正确地、恰如其分地运用并不容易,需要结合研究对象的特点。

郑真真副教授从另一个角度对描述性统计、双变量统计和多变量统计分析中经常被忽视的问题进行了讨论和分析。她不仅指出了问题的所在,而且告诉大家应该如何做,她所讨论的内容是大部分统计教科书上所没有的,具有很大的实用性,对于正确地应用统计分析方法具有很大的指导意义,具体的问题这里就没有必要再评述了。她的分析源于多年教学和科研工作的总结和升华,很多观点具有独到性,如对描述性统计的重视,多变量分析结果的展示和解释,不必求最新、只求最合适的方法。

王谦和郑真真从不同角度对统计分析和统计数据应用中出现的问题进行了分析,郭志刚教授则对人口统计中的一个重要指标—总和生育率进行了深入分析。他指出指标是否有效应该以研究目的为标准来判断,他从总和生育率对终身生育水平的背离和不能控制育龄妇女的孩次结构这两大缺陷出发,对问题的表现形式、实际工作中的需求以及改进方法进行了分析和介绍,其中渗透了他自己近年来对这一问题的研究结果和体会。我看完他的文章觉得受益匪浅,如 90 年代以来中国的人口调查所得到的总和生育率指标都非常低,这里面有统计漏报的问题,也有总和生育率指标自身的缺陷所带来的问题,在婚育年龄推迟时,总和生育率会低估终身生育水平。在以往的人口学研究中,对方法论的讨论和重视是不够的。研究方法的创新源于实践的需求,面对 21 世纪更加错综复杂的人口问题,我们需要研究方法的不断创新与发展,这是目前我国人口科学发展面临的一个巨大挑战。

三位专家的分析 and 讨论有很大的实用性,值得从事人口和计划生育的实际工作者和学者认真学习,但是毕竟他们不可能把所有的问题都归纳概括全面,这也不是本期论坛的初衷。我们希望通过本期论坛起到抛砖引玉的作用,引起同行对这一问题的重视。中国规范的人口学研究毕竟只有 20 多年的历史,有问题是可以理解的,统计分析方法的误用和研究的不规范在其他学科领域也同样存在,甚至有可能更严重,但是我们不能漠视这种状况继续下去,问题的解决需要全体同仁的重视和努力。我觉得目前至少可以加强以下几个方面的工作:

- (1) 提倡学术争鸣,鼓励同行之间开展批评,这将有利于学科的发展与规范;
- (2) 编审人员把好质量关,使在杂志上发表的文章符合学术规范,杂志上文章的示范作用往往比教科书要大得多;
- (3) 加强统计分析方法的教学工作。强调方法的适用性和在实际应用中应该注意的问题。

(主持人 刘金塘)