

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决定系数较低时,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来进行。

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。两种性质都反映在各年龄组生育率的合计上。通常,不同基

在这一方面,总和生育率与人口的粗再生产率和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。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,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。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。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。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。

如男性平均年龄加女性平均年龄没有意义。

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”。

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。

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。

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。

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。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。

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(MAG)也会变化。MAG 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 MAG 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延对当前生育水平影响的估计。

我认为,尽管 TFR 指标还有继续改进的余地,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。

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

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。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。

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。

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。

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。

比如某一年的 MAG 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,调整的结果也并不差。

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。

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。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。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。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。

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

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。

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。

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)。

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。

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获益匪浅,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)