

编者按 人口学从诞生的那一天起,就不断地从其他学科吸取营养。近年来,人口学与相关学科的交叉研究方兴未艾,这不仅拓宽了人口学的研究领域,而且使人口学者观察问题的视角更加宽广、多样,同时对人口科学研究的深化起到了积极的推动作用。这里,我们编辑了一组介绍交叉研究方法的文章,以飨读者。

历时研究与事件史分析

郭志刚

【提要】 本文分析了截面研究与历时研究对于因果关系研究的有效性,论述了不同类型的历时研究,提出了本源性变化的新概念,并对事件史分析进行了简要的介绍。

【作者】 郭志刚 北京大学社会学系,教授。

科学研究旨在探索事物发展的规律,其中因果关系研究是重要的内容之一。因果关系研究是解释性的研究,至少要涉及两个变量(一个作为因变量,一个作为自变量)。由于社会科学研究难以像自然科学研究那样将研究对象置于实验室中高度控制的研究条件之下,又因为社会变量之间也存在着错综复杂的相互影响,而且作为社会科学主要研究对象的人对于实验会产生心理和行为反应,从而损害实验结果的有效性,所以,社会科学的量化研究更多采用调查分析的方法。

一、截面研究与历时研究

调查分析可以分为截面研究和历时研究两大类(艾尔·巴比,2000)^①。这两种类型的划分着眼于数据的时间维度。目前国内社会科学的量化研究主要采用截面研究。这种研究的特点是进行一次性调查,取得一个时点上不同分析单位的数据来进行研究。比如,人们常比喻人口普查是对全国人口状况进行照相。就一次普查数据来分析,就是截面研究。这种数据对静止状况的描述性研究更有效(如反映全国人口在普查时点的数量与结构),而对于解释因果关系的有效性较差。从研究方法论而言,判断因果关系是否存在有三个必要条件(艾尔·巴比,2000:98):(1)在时序上先有因后有果;(2)两个变量之间具有实证相关性;(3)两个变量间的实证相关不因为第三个变量的存在而存在。本文要讨论的是第一个必要条件时序。由于截面数据只能反映时点差异,不反映时序变化,因而使得这种数据不利于因果关系研究。从方法论角度讲,差异指同一时点上不同对象之间的差别,因而不涉及时序;而变化则

^① 过去关于研究方法的文献中常分别称之为横向研究和纵向研究。这里沿用了邱泽奇在《社会研究方法》一书中的译法,因为其译义更为明确。

指同一对象在不同时点上的差别,因而必然涉及时间。换句话说,只有历时性数据才能反映真正的变化,也才有可能较为直接地探讨因果关系问题。

虽然用截面数据进行因果关系研究已经司空见惯,其实这种研究是在将不同对象之间的差异当作同一对象的时序变化,其中必然隐含了一个重要的假定,即所有对象是按同一模式发展变化。比如,在采用截面数据进行的生育水平的影响因素研究中,实际上要假定,发达国家的今天代表着不发达国家的明天,而不发达国家的今天代表着发达国家的昨天。尽管有时这种假定比较接近于实际情况,有助于推进研究和增进认识,但这种假定并不总是符合实际的。这种以经济发达国家为一极、经济不发达国家为另一级的发展理论经常遭到反对,被认为是西方文化中心论。事实上人们看到文化之间的差异并不随经济发展而消除(如日本与美国),这反映出先发达地区的情况未必是后发达地区未来的归宿;另一方面,也有研究(李中清、王丰,2000)指出,中国人以前乃至现在的人口行为与西方人过去的经历十分不同^①。这类怀疑和批评其实直指将截面数据作为历时数据来用时,假定条件背后的理论基础。

简而言之,仅仅依据截面数据并不能证明案例之间的差异并非先天存在,也不能断定案例的不同状况真的经历过什么变化,因此再讨论“变化”的原因是很牵强的。而对于纯粹的差异来说,案例类别是先天决定、不能改变的,其状态水平只是类别的一种属性,类别与属性之间的关系并不是真正的因果关系,即自变量和因变量都不能变。

近年来,国内社会学界对于历时研究越来越重视,生命历程、生命周期、同龄群体的概念被引入作为研究对象或作为研究方法(李强、邓建伟、晓箏,1999),这些实际上都是历时研究的范畴,因而也是更有效的因果关系研究的方法。

二、历时研究的三种类型

艾尔·巴比(2000)将历时研究归纳为三种类型,即趋势研究、队列研究、固定样本研究^②。归纳巴比对每一种方法的具体介绍,笔者认为三种类型的本质不同在于它们的分析比较单位不同。

趋势研究的分析单位是相同特征群体,研究其历时变化。它的对象既可以为总人口,也可以是某一性别年龄组。通过比较来把握同一口径群体的变化趋势。由于其统计数据较易获得,因此在实际中应用普遍。由于趋势研究的分析单位只是一种统计口径,因此所抓住的变化是十分笼统的,既可能是分析单位内部构成的变化,也可能是部分对应个体所发生的本源变化。比如,两次普查所得到的20岁组中的个体完全是不相同的人,因此所说的趋势变化完全是统计上的变化,而没有本源变化。又如2000年中国受教育水平比1990年的提高不一定是因为1990年的那些成年人的受教育水平真的提高了,而可能是因为两次普查期间加入成年人口行列的新一代比老一代受教育的水平提高了,而受教育水平较低的老一代中许多人已经退出了这一人口行列。也就是说,趋势研究旨在描述某一口径的统计变化趋势,并不在意变化的来源和因果关系。

队列研究的概念对于人口学研究非但不陌生,而且它正是人口统计学的特色之一。人口学的队列研究主要是研究整个队列的终身情况,比如终身生育水平或生存寿命及其相应的模式。由于人类生育期和寿命都很长,人口学的队列研究常常等不及终身数据的取得,更多地采用了时期年龄别指标,进行假设队列研究(如时期预期寿命和总和生育率)。人口预测则进一步假定基期以后各个队列

① 李中清、王丰挑战由马尔萨斯首次提出、并经过其他学者改进的人口行为模型,揭示出“东西方之间的人口行为存在着的很多差异,而这些差异是社会组织的地区和历史差异”。

② 邱泽奇的中译本将后两种类型译为“世代研究”和“小样本多次访问研究法”。这里沿用吴辉主编的《英汉统计词汇》(1987)中的译法。

的生育和死亡的年龄别变化率按照同一水平和模式进行历时的变化。其实,真正的队列研究侧重的是队列之间不同的群体行为特征,并力图将其归因于共同经历某些历史事件所形成的队列效应或由于与前后队列相比规模很大或很小所带来的队列效应。但是假设队列法却旨在构造具有一般意义的队列终身水平和年龄别模式,这与本来意义上的队列研究有所不同。正是由于这种人口学研究已经成为一种规范,其内在的缺陷往往被人们忽视和遗忘,或者只注意到未来水平的变化,忽视了模式的变化。

既然是假设队列研究,因此需要不断地在一定时间后对部分真正队列的年龄别变化(也是真正的历时变化)进行比较和再检验。比如,在用人口普查等截面数据资料对人口居住方式的年龄别分布研究中(郭志刚,1995),曾发现在老年段中生活在三代以上户的比例随年龄显著提高的“变化”。其实,这里可能存在将不同队列迭加的效应误当成同一队列历时年龄效应的危险。而直接对1982年各队列在1990年的情况进行了追踪比较后(杜鹃,1999),才能够肯定存在队列历时性(即年龄上的)变化,而不是时点上的队列与年龄的混合差异。在人口学研究中,年龄效应、时期效应和队列效应往往搅在一起,80年代兴起的APC模型^①是试图分解各因素的统计方法(彭非、WANNER,2000)。

需要说明,队列分析虽然追踪了某个“批”,但并不需要追踪这个批中的每个案例。比如,一个老年队列在前次普查时的个体中已经有相当一部分因死亡而不包括在后一次普查的同一队列中,这并不影响作为集合的队列分析。但是,如果队列人口的死亡(或其他原因的丧失)有很强的选择性时,比如生活在不同类型户中老年人死亡率差别很大,那么队列分析将损失对本源性变化的把握,这一历时变化将可能只是由于失去某种特征的个体而发生集合之中的结构性改变,而留存的个体则可能实际上没有变化。由此看出,这与趋势研究有类似的问题,即只要分析单位不是真正承担变化的个体,汇总的“变化”并不是本源意义上的变化。

固定样本研究便是旨在观察这种本源变化的方法。它要求在一定期限内追踪样本中每一案例,多次记录其在不同时间点的情况。固定样本通常是要追踪承担本源变化的个体,以达到更好地研究因果关系的目的。比如,尽管我们倡导老年人采取健康的生活方式,而老年人由于生理功能的退化,队列中健康者的比例不可避免地出现下降。那么老年人采取健康的生活方式究竟有没有用呢?北京市老龄化多维纵向调查便是一项固定样本的历时研究(项曼君、孟琛、汤哲,1997),其结果揭示出虽然从群体上健康状况的退化依然存在,但从个体上并不是与此一致,采取健康生活方式的老年人更有利于缓解健康水平的退化,有些个体甚至有所好转,而采取不健康生活方式则会加速个体健康水平的退化,从而以实证量化分析肯定了采取健康生活方式的必要性。

比较三种历时研究,可以看到一种梯次,分析单位越接近本源变化的直接承担者,对于研究因果关系的有效性就越强。尽管历时数据对因果关系的揭示更有效,但由于其收集成本高、技术难度大,限制了历时研究的应用。

三、事件史分析方法

限制应用历时研究的不光是历时数据的收集难度,其实还有历时数据的分析难度问题。很长时间以来历时研究只能进行汇总水平变化的比较分析,无法建立多元解释变量的综合分析模型。近年来,事件史分析方法在社会科学研究领域中异军突起,越来越流行。

根据Yamaguchi(1991)的定义,事件史分析专门研究“事件发生的方式及其相关因素”。事件史分析在不同学科有不同的称谓,在生物医学研究领域(如癌症患者存活时间),它被称为“生存分

^① APC模型指同时对年龄效应、时期效应和队列效应进行研究的模型。

析”，而在工程研究领域(如灯泡的寿命)，又被称为“失败分析”。

事件发生的方式可以用很多形式表达，比如风险函数、生存函数、概率密度函数、累计分布函数。它们都是事件发生方式的时间函数。其实，从数学上来说这四个概念是等价的。也就是说，要是我们知道这四个函数中的一个，就可以推导出其他所有函数。然而所有的事件史分析的统计模型都倾向于使用风险函数，而不用其他函数，这是因为其使用方便和易于解释。

人口学中发展起来的生命表可能是历时研究中最古老而又相对复杂并完备的事件发生时间分布的分析方法。生命表中的死亡概率 q_x 便是一种风险函数。而生存函数反映死亡事件发生于某一时点之后的概率，即为各时点的尚存人口占原有人口的比例(即： $S_x = l_x/l_0 = (\sum_{i=x}^{\omega-1} d_i) / l_0$)。至于累计分布函数则相反，反映了人口中死亡事件发生于某一时点之前的概率，即： $F_x = 1 - S_x = (1_0 - l_x) / l_0 = (\sum_{i=0}^{x-1} d_i) / l_0$)。注意：生命表以时间分布形式描述群体死亡事件的发生方式，然而对每一个体而言只有死亡的发生时间问题，更确切地说是其出生至死亡的时间间隔问题。事件史分析方法定义这种间隔为持续期，只要我们换用另外一对起点和终点，便可以将这种方法推广到更广阔的研究领域，比如从结婚到初育、从参加工作到初次晋升等。事件史分析的另一个基本概念是风险集，它表示一批在不同时点上经历某一事件风险的对象，与生命表中的尚存人口相对应。

然而生命表采用汇总形式统计，损失了个体的其他特征信息与死亡年龄之间的联系。如果要研究死亡与其他影响因素的关系，只能先将人口进行分类(如按性别)后，分别计算生命表并比较相应的统计值。这种方法效率很低，需要很大样本规模才行。比如，当我们只考虑性别和城乡两个最简单的分类变量，便需要计算四个生命表来加以对比。当变量为收入或受教育年限等连续变量时，便先要将其转换为只含少数类别的分类变量，否则根本不可能再用生命表比较的方法了。

至于像回归分析这样的常规多元统计方法并不能应用于事件发生时间的研究，是因为事件史数据有两个基本特征：(1)风险或生存在时间上都属于倾斜分布^①。这样一来，便违反了回归方法要求变量是正态分布的假设条件。(2)事件史数据存在大量删截现象。所谓删截指调查时大量案例尚未发生死亡(或其他研究事件)，即生存尚在继续，其死亡时间未知^②。常规的回归模型不能处理删截案例。而且，将删截案例排除于分析之外也不行。一方面，这样做便浪费了信息，因为删截案例虽然没有提供确切的死亡发生时间，但并不是没有提供信息，它们提供的信息是到调查时它们已经生存了多少年。另一方面，要是删截案例与未删截案例之间存在系统性差别，如删截案例普遍教育水平较高，那么排除删截案例后的教育水平影响的回归结果便会产生系统性偏差。

此外，常规回归不能容纳动态自变量(也称时变变量，指随时间而变的自变量)。在历时研究中，一类自变量只因人而异，但是不随时间而变(如性别、种族)；另一类自变量不仅因人而异，而且随时间变化。这种动态信息并不能通过在模型中纳入多个代表不同时点的自变量来解决。例如，婚前不久的收入对于是否结婚有特殊重要意义，即这时的收入与结婚有具体时序间隔要求，这时在模型中纳入任何确定年龄的收入都不解决问题，因为每个人的结婚年龄不同。如果确定了较小的收入参照年龄会导致部分对象在这个年龄离结婚还很远，而确定较大的参照年龄则会使另一部分对象是在用婚后的收入来解释是否结婚，违反了前因后果的原则。

事件史分析可以令人满意地解决变量的偏态分布、删截及动态自变量等历时数据中特有的问

① 熟悉生命表的人都知道死亡概率为J型曲线，除婴幼儿期有所下降外，之后随年龄单调上升；而生存曲线(即 l_x 曲线)则是单调下降的。

② 实际数据中删截的情况十分复杂，这里只介绍最为常见的一种。

题,因而可以充分利用其中所包含的相关事件明确的时间顺序,将因果关系定量地揭示出来。虽然我们经常形象地比喻事件史分析是“生命表+回归分析”,但通过以上比较,可以看到其实它并不是这两种方法的简单相加。

事件史分析模型分为两类,即离散时间模型与连续时间模型。这涉及到持续期的时间单位,隐含着事件发生时间的测量准确性问题。时间本来是连续变量,但如果时间单位取得很大(如以一年或更长的时间),通常将其作为离散时间对待。

对于离散时间单位,主要采用离散时间 Logit 模型。其思路是: $P(t)$ 代表某人在时间 t 上发生某事的概率。我们可以运用下列 Logit 模型方程简化式拟合观察数据:

$$\ln \frac{P(t)}{1-P(t)} = a(t) + b_1 x_1 + b_2 x_2(t)$$

其中, x_1 代表一般自变量, $x_2(t)$ 代表动态自变量。 $a(t)$ 是截距,只要模型中还有代表不同时期的虚拟变量,它便会随时期变化。这样,我们就可以应用常规 logistic 回归进行系数估计和检验,解释各自变量对事件的影响。

在分析连续时间单位的事件史数据时,又有非参数模型和参数模型之分。非参数模型用不着对事件发生分布作任何假设,而参数模型则需要假设事件发生分布服从某种数学函数。当理论上已知事件发生时间分布时,非参数模型不如参数模型有效,因为数据所包含的信息不能得到充分利用。然而,在事件发生时间分布未知时,非参数模型却更有效。在这种情况下,因为非参数模型更为稳健,因而可以避免由于错误设定事件分布可能造成的误差。

在很多情况下,事件发生时间的参数形式是未知的,需要对此加以估计。在大样本时通常采用生命表估计,将持续期的时间长度划分为所需的间隔分组来计算事件发生的时间分布。在小样本时通常采用 Kaplan—Meier 估计,它实际上是生命表的一个特例。这两种方法都属于非参数估计方法,即用不着对理论分布作任何假设。另外,还可以通过参数方法来进行估计。

连续时间模型有很多,但使用最广的是 Cox 比例风险模型。用 $h(t)$ 来代表风险率,Cox(1972)提出可以估算下列比例风险模型:

$$\ln h(t) = a(t) + b_1 x_1 + b_2 x_2$$

这里 $a(t)$ 是基准风险函数,可以以任何形式出现。由于 $a(t)$ 并未规定,从这个意义而言,Cox 模型是一种半参数模型,根据对 $a(t)$ 的不同规定,它可以转化为非参数模型或不同的参数模型。并且,由于对任意两个案例 i 和 j ,两者的风险之比是一个常数,只取决于个人特征而与时间 t 无关。从这个意义上讲,Cox 模型又是成比例的。然而,一旦引入动态自变量(比如将 x_2 换为 $x_2(t)$),这个模型就不再是成比例的了。Cox 的巨大贡献在于对这一模型的估计提出了很好的方法。

在某些研究场合中,还需要对事件风险的时间分布提出假设,比如,指数分布模型、Gompertz 分布模型、Weibull 分布模型。这三种模型都属于比例风险模型的大类,并且都可以通过最大似然估计方法来估计参数。但是,这三种模型的一个缺陷是,在加上动态协变量以后模型估计就变得很难。此外,还有一种加速失效时间模型。

至于哪一种模型最好,依赖于不同的研究需要。并且在考虑拟合模型的类型时,还需要知道所用的计算机软件能够拟合哪些模型。社会科学研究经常应用的估计生存模型的常用软件有:SPSS、SAS、STATA 等。其中 SPSS 菜单化程度很高,利用其 logistic 回归功能,可以估计离散时间 logit 模型。在其生存分析部分包括有:生命表估计、Kaplan—Meier 估计、Cox 比例风险模型(分为不带动态协变量的和带有动态协变量的两种)。但它还不能估计其他参数模型。SAS 软件和 STATA 软件的统计功能比 SPSS 软件强大,如 SAS 在其 PROC LIFEREG 程序还可以做加速失效时间模型分析,STATA 可以进行许多关于生存模型的统计检验,还可以估计 Weibull 和指数模型。

事件史分析方法对于数据有特殊的要求。它必须采用固定样本的历时数据。通过追踪调查得到的固定样本数据是最理想,但这种调查要求事先有极为细致的研究设计,一旦开始调查项目便不能更改。另外,通过在一次性调查中加入回顾性问题也可以得到事件史数据,如中国多次生育率调查中都包括生育史的信息。但严格地讲,这种数据并不是事件史分析最理想的数据,因为虽然生育变量具有历时动态,然而作为解释变量的其他特征(如受教育、城乡等状况)则是调查实施时的截面状况,而不是事件发生时的状况。在使用这些解释变量时实际上要假设其值从研究事件发生前至调查时间没有改变过。这种假设有时比较合理,有时则不一定合理。段成荣(2000)的研究表明,在是否迁移的常规 logistic 回归中,尽量根据调查提供的信息将解释变量值“倒推”回迁移之前的状况,能够显著提高模型的解释能力和各解释变量作用的显著程度。另外,还需要注意回顾性调查往往会发生遗忘性、隐瞒性误差,并且还可能导致死亡选择性带来的系统偏差。

此外,事件史分析有比较特殊的数据格式要求,在正式开展分析之前往往需要对事件史数据进行预先处理。

总之,通过以上评述可以看到,掌握历时研究的有关概念对开展研究工作具有很强的方法论指导意义。至于实际应用历时研究及事件史分析则要求有更长期的研究计划、更大的科研投入、更高的方法论素质、更强的技术水平和操作能力。否则,损失也将更为严重。因此,应用历时研究切忌仓促上阵与简单化、形式化的粗制滥造。单纯应用某一两种研究方法和分析技术还不完全等于研究的精致与创新,并不代表科学研究已经达到很高的水平,因为科学研究工作需要综合素质与能力的提高。

参考文献:

1. 艾尔·巴比:《社会研究方法》(上、下)。邱泽奇译,华夏出版社,2000年。
2. 李强、邓建伟、晓箬:《社会变迁与个人发展:生命历程研究的范式与方法》,《社会学研究》,1999年第6期。
3. 李中清、王丰:《人类的四分之一:马尔萨斯的神话与中国的现实,1700~2000》,三联书店,2000年。
4. 吴辉主编:《英汉统计词汇》,中国统计出版社,1987年。
5. 内森·凯菲茨:《家庭人口学的形式和实质》,载于邦加茨、伯奇、沃克特编:《家庭人口学:模型及应用》,曾毅等译,北京大学出版社,1994年,第一章。
6. 郭志刚:《当代中国人口发展与家庭户的变迁》,中国人民大学出版社,1995年。
7. 杜鹏:《中国老年人居住方式变化的队列分析》,《中国人口科学》,1999年第3期。
8. 项曼君、孟琛、汤哲:《1992~1994年北京市老年人生活变化的调整》,《人口研究》,1997年第3期。
9. 彭非、WANNER:《APC模型及其对中国生育率数据的分析》,《统计学》,中国人民大学书报资料中心,2000年第1期。
10. 段成荣:《影响我国省际人口迁移的个人特征分析——兼论“时间”因素在人口迁移研究中的重要性》,《人口研究》,2000年第4期。
11. Cox, David R. (1972), "Regression Models and Life Tables." *Journal of Royal Statistical Society*, B34: 187-220.
12. Yamaguchi, Kazuo. (1991), *Event History Analysis. Applied Social Science Research Methods Series Volume 28*. Newbury Park, CA: Sage Publications, Inc.

(本文责任编辑:朱 萍)