

升学目标地数据研究教育机会 不平等的局限性

——以“首都大学生成长追踪调查”为例

李 代

内容提要:在升学过程中出现的教育机会分配不平等问题是社会分层领域的重要研究课题。一些定量研究采用的数据是升学目标地调查数据,也就是从完成升学的大学生人群中采集的调查数据。本文以“首都大学生成长追踪调查”的数据生成过程为例,将模拟生成数据与某省完整高考数据进行对比,基于“阈值依赖不平等”的研究框架,试图说明基于目标地数据得到的定量研究结果可能在可阐释性和可推广性方面存在局限,而且“首都调查”自身在无偏性方面存在缺陷,可能导致有误导性的结果。

关键词:教育不平等;高考;社会分层;社会流动

从知识进化论(波普尔,1987)的视角来看,学术研究往往是在已有基础上的积累前进,当理论、方法或数据当中任何一方面取得进步时,新的研究成果得以产生,而此前被学术界接受的研究结论就有可能被置于怀疑的境地。一种常见的情况是原有研究者受限于现实条件而只能抛弃理想化的研究方案,转而采用可行性更强的方案。这样的条件下取得的研究成果尽管未必能令人完全信服,但向可行性的妥协造成了多大的偏差难以评估,在没有更好的选择时也有可能说服学术界加以接受。一旦新的研究者有条件采用更为严谨的研究方案对同一问题进行研究,此前的研究结果就可能需要得到更新。例如,由于无法观察到总体情况,研究者采用抽样调查来获得对总体参数的无偏估计;由于数据采集成本过高,研究者实施的抽样方案未必能满足随机抽样的理想条件,而会存在一定妥协。这些妥协往往伴随着一定的代价,但是由于抽样调查的稀缺,针对同类问题往往只有屈指可数的可比较的调查数据,因而基于这些数据进行的研究缺乏有效的反馈回路来校验其结果的可靠性。

本文试图提供对教育不平等议题下一类研究的校验。教育不平等本身是非常重要的研究课题,获得对总体参数的准确估计不论是对学术研究还是政策制定都有非常明显的价值,因此对其进行精细的讨论很有意义。教育不平等有多种表现形式,而得到研究最广泛的问题之一便是不同群体在获得教育机会方

作者简介:李代,北京大学新媒体研究院助理研究员,主要研究方向为社会分层与流动。本文感谢周飞舟、刘亚洲及匿名人的重要建议。

基金项目:中国博士后科学基金资助项目“基于大数据视角对高等教育机会不平等的研究”(项目编号:8206300008)。

面的不平等。例如,家庭背景、地域、城乡、性别、民族等等重要的变量都可能对教育机会的获取产生影响,从而造成群体之间的教育不平等。到底不同群体之间的教育不平等差异多大?不平等的变化趋势如何?这一差异是如何产生的?

要准确地回答这些关于教育不平等的问题,定量研究不可或缺。对教育不平等的定量研究,采用的数据可以分为三类。第一类是一般性的调查数据,例如中国1%人口抽样调查、中国儿童情况抽样调查(张春泥、谢宇,2017)、中国家庭追踪调查(陈伟、乌尼日其格,2016;李忠路、邱泽奇,2016)、中国综合社会调查(杨中超,2016)等等。第二类是生源地数据,这类数据的研究总体是参与同一场教育竞争的人群。例如,某年某省全部参与高考的学生构成的数据,就是一个典型的生源地数据(李代,2017)。与之相对,第三类是目标地数据。在中国,来自全国不同生源地的学生经过高考而汇聚到大学之中。他们所进入的大学,便是我们所说的升学“目标地”,而从大学生中获得的数据,便构成了目标地数据。例如,首都大学生成长跟踪调查(吴晓刚,2016)、首都高校生发展状况调查(叶晓阳、丁延庆,2015)、学籍卡数据(梁晨等,2012;梁晨、董浩,2015)。

这三类数据的本质差别不在于是否进行抽样,而在于总体的构建与作为研究对象的升学机制是否相互独立。一般性调查数据在构建总体时并不涉及升学的问题,也就是说总体的构建与升学相独立。生源地数据在构建总体时,选取的是升学发生之前就能定义出来的、相互竞争教育机会的人群。目标地数据在构建总体时,选取的是升学过程产生的、获得了教育机会的人群。从理论上讲,要测量升学过程中出现的教育不平等,前两种数据都可能提供准确的结果,而第三种数据多数情况下不能提供准确的结果,因为它缺失了教育竞争中失败者的信息。尽管有不可回避的缺陷,但是因为数据收集方面的便利,目标地数据在大量的研究中得到了使用。本文想指出的是,采用目标地数据进行定量研究,得到的结果很可能存在可阐释性方面的问题、可推广性方面的问题,这些都使得研究结果的价值大打折扣,而且难以与前两类研究得到的结果直接进行比较。

为了说明这些问题,本文以“首都大学生成长跟踪调查”(后文简称“首都调查”)的数据生成过程为例进行分析。之所以选取这一数据进行案例分析,是因为它是典型的目标地数据,而且有不少重要的研究都基于这一数据进行,例如,吴晓刚(2016)、谢桂华和张阳阳(2016)、李骏(2016a, b)、李忠路(2016)、许多多(2017)、朱斌(2018)等等。由此可见,这一数据在研究教育不平等议题方面是质量相对较高的,否则不应产出这么多得以在优秀学术期刊发表的成果。以这一数据为例进行讨论,有更大的典型性。这一调查数据并未公开,因此本文并没能掌握数据本身^①。但是数据生成的过程在研究者的文章中得到清晰的介绍(吴晓刚,2016;李路路,2013),这使我们可以对某一总体数据进行筛选,按照“首都调查”实施抽样的步骤生成模拟数据,从而判断是否能够基于该调查数据得出可靠的研究结论。

要生成模拟数据,本文使用的是某省某年的完整高考数据。该高考数据中包含了全部报考考生的个人信息、考试成绩及录取结果。这一数据包含了全部报名考试学生的信息,因此是一个总体数据。由于包含了录取结果,对其按一定的条件取子集,能够生成目标地数据。如果目标地数据是有效、可信的,那么用总体数据和模拟生成出来的目标地数据分别进行同样的分析,得到的结果应该是完全一致或至少非常接近的。如果不然,就说明用目标地数据进行分析的结果不那么可靠。因此,对生成的目标地数据重复对总

① 笔者开展研究时,该项目基期调查的数据已经从中国人民大学中国调查与数据中心的网站上下线了。

体数据进行的分析,比较二者结果的差异,可以评估目标地数据的可靠程度。

本文按照“首都调查”的说明,从完整的数据中生成目标地数据,穷举分数线从1分取到700分的情况,分别计算相应的性别不平等程度。之所以选择性别不平等作为研究的对象,出于以下三个考虑。第一,抽样调查数据对研究者的允诺是提供对总体参数的无偏估计,因而任何变量上都不应该出现较大的偏误。所以,本文不需要对所有变量进行详尽的分析,只要举出一个反例就足以说明问题。第二,选择性别变量作为研究的焦点,是因为性别在社会学研究中广受关注——前面列举的研究无一例外都控制了性别变量;在研究人群中性别比例比较均衡、分数分布也存在差异,描述性别不平等比较容易揭示出问题。第三,之所以选取李代(2017)中“阈值依赖不平等”研究框架作为重复研究的内容,是因为这一研究框架在每一个可能的分数线处计算一次升学差异,例如从1分到700分计算700次假想的分数线带来的结果,就相当于进行了700次逻辑斯蒂回归,得到的结果可以带来对整体情况全局的把握。这比起以一本线、二本线等标准计算一次回归提供了更多的信息,能更好地展现不同人群升学机会差异的结构。

本文的研究结果包括以下几方面内容。首先,作为一个抽样调查,“首都调查”面对所有抽样调查都面对的挑战。由于中国的高考是分省、分文理进行的,如果在调查研究和模型设置时不对这些变量加以控制,就可能导致一类“生态谬误”——辛普森悖论(Simpson's Paradox)(Simpson, 1951)。吴晓刚(2016)试图通过将考试分数标准化来解决这一问题,但其效果并不理想。如果控制省份、文理的变量,就可能导致样本量不能满足模型要求,从而得不出显著结果。其次,“首都调查”作为一个目标地抽样调查,存在两类特定问题。由于其总体是升学的结果,难以放入日常生活经验中加以理解,因此有可阐释性问题;要想通过把结论推广到可阐释的人群上去来解决可阐释性的问题,又发现其结果不具备很强的可推广性。最后,就“首都调查”自身而言,因为采用的抽样方案实施效果不甚理想,样本对研究总体的参数估计很可能是有偏的。

本文是一篇反思性研究,提出的问题是笔者在研究中也经常面对而没有良好解决方案的难题。但是另一方面,如果本文能提醒研究者意识到这类研究的缺陷、明确其研究贡献的限度,甚至激发学者提出有效的解决方案,本文也就不失为有微末的贡献了。

一、“首都调查”的调查设计

吴晓刚(2016)对“首都大学生成长追踪调查”(Beijing College Students Panel Survey, BCSPS)进行了详细的介绍,下面首先对其中与本文相关的部分加以总结,给后面的论述提供背景。该调查研究的总体是“北京市范围内所有教育部直属和其他中央部委或北京市所属非民办大学的全日制本科生”。项目研究的学生是2006年与2008年秋季入学的本科生,抽样框是当时北京市54所公立学校2006级和2008级在校学生学籍数据库。抽样采用多阶段、分层、概率与规模成比例(PPS)的抽样方法。研究者把54所大学划分成6层,其中北京大学、清华大学和中国人民大学单独分层(层1、2、3),剩下的学校里所有教育部直属或者其他中央部委所属的“211”大学为一层(层4),所有教育部直属或者其他中央部委所属的非“211”大学为一层(层5),北京市属地方高校为一层(层6)。各个高校应该划分到哪一层中在多数情况下都很清楚,但是有三所大学存在模糊性。中国矿业大学(北京)、中国石油大学(北京)和中国地质大学(北京)分别是三所大学的北京校区。尽管这三所大学本身都是“211”大学,但是北京校区不一定出现在“211”大学名单上,例

如中国矿业大学(北京)被划成了非“211”大学。因为不清楚另外两所大学被如何对待,本文依照中国矿业大学(北京)的例子,假设中国石油大学(北京)和中国地质大学(北京)也被当成非“211”大学。

调查采用PPS抽样法,从北京大学、清华大学和中国人民大学各抽取了500名学生,以确保顶尖学生的数据能被收集到。对其他3层的学校,先根据各层学生的数目成比例抽取学校,分别抽到6、2、4所高校;再从每个高校抽取300人。这样得出了15所高校、5100人的学生样本。2009年进行的基期调查完成4771人,之后还对这些学生进行了追踪调查。在数据清理的过程中,研究者提供了各校学生的权重,希望能把定量研究的结果推广到总体上去。

根据吴晓刚(2016)的研究模型和李路路(2013)对基期调查数据的描述,可以知道数据采集了哪些与教育不平等相关的变量。首先,大学分为精英大学、“211”大学和一般院校。高考分数进行了标准化,标准化的方式是:考生标准分=(考生考分-本省一本线)/(本省一本线-本省二本线)。学生的生源地不是按省划分的,而是被分为四个层级,即“农村与乡镇”“县级市”“地级市”和“省会城市或北京”。上大学前的户籍、性别、民族等变量也包含在调查中。基于这些变量,吴晓刚(2016)把家庭经济地位、家庭居住地、就读高中类型、特殊录取待遇、标准化高考分数、性别、民族、入学年份等放入模型,运行了若干线性回归模型和逻辑斯蒂回归模型,来考察这些变量对升入不同类型高校的影响。

在中国,高考是分省、分文理进行的,招生录取同样如此。这就意味着如果大学在各省、文理招生的名额数量不同、考生数量不一,那么录取的难度差异可能相差悬殊。因此,不控制省份、文理而谈升学,可能会带来一定的偏差。吴晓刚(2016)考虑到了地域差异的问题而对高考分数进行了标准化,那么其标准化的方案是否能解决这个问题呢?

二、作为抽样调查的一般性问题

(一)标准化能不能消除省份、文理差异

吴晓刚(2016)对高考分数进行了标准化,标准化的公式为考生标准分 = (考生考分 - 本省一本线) / (本省一本线 - 本省二本线)。显然,如果考生分数恰为一本线分数,那么在标准化之后其分数为0;如果考生分数恰为二本线分数,那么在标准化之后其分数为-1。这一标准化的效果是让不同省份、文理之间一本线、二本线之间的差异消失。但是如果研究的因变量是“是否能升入精英大学”“是否能升入‘211’大学”,这一标准化是否起到其应有的作用呢?下面以北京大学(含北京大学医学部)、清华大学和中国人民大学构成的“精英高校”为例,来考察标准化是不是有效。

笔者没有其他省份的高考数据来进行比较,无法直接判断这一标准化的方案是否能消除省份之间的差异。但是可想而知,如果标准化的方案是有效的,那么应该也可以消除同一省份文理录取的差异。如何判断差异是否被消除呢?如果标准化之后精英大学的文、理分数线重合了,说明标准分高于分数线的文、理考生都能被录取,而低于分数线的文、理考生都不能被录取,文、理分数的差异也就被消除了。

如何得到这一分数线呢?现实生活中的分数线是被某高校常规录取的投档分最低考生的投档分。对该省当年三校分数线的查询发现,经过标准化之后文科分数线为3.35,理科为5.24,相差甚远。由于不确定性因素,当考虑的高校较少时分数线可能会被极端值拉低,因此可以采用另一种方式对其进行定义。在理想情况下,如果三所学校在该省理科招生总数为 n 人,那么该省的前 n 个理科生都应能成功被这些学校

录取。此时,第 n 名理科生的投档分就可以被看作理科分数线。同理,可以对文科也计算分数线。下图是标准化前、后的投档分分布与精英大学分数线位置,文科用虚线表示,理科用实线表示。

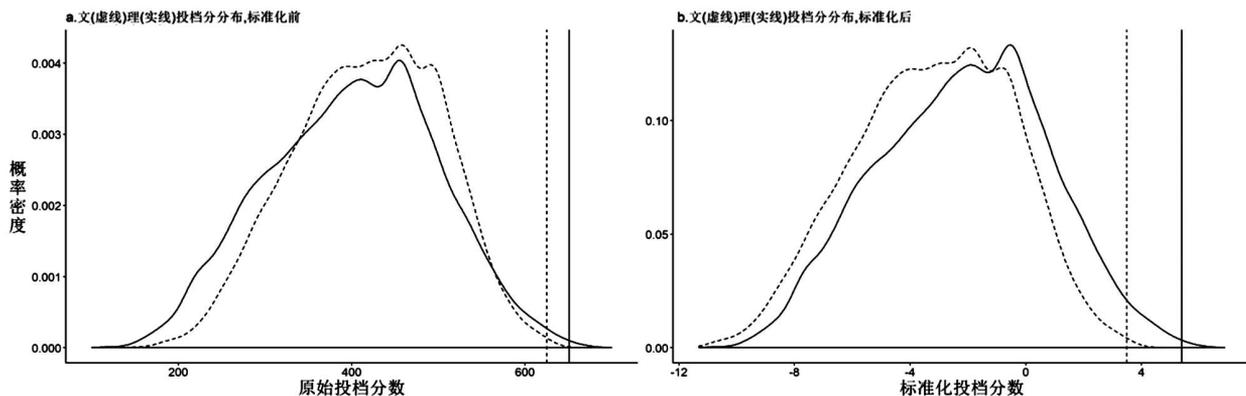


图1 标准化对精英大学文、理分数线的效果

从图1中可以看出,文、理分数线没有重合。如果考生的投档分在标准化之后落在两条分数线之间,那么同样的分数下文科生能够被录取,而理科生不能被录取。这说明这一标准化方案并没有消除文、理科高考投档分作用的差异。既然如此,想必它同样难以将分数的地域差异消除。因此,把模型中的分数变量标准化仍然不能排除地域、文理科的影响。那么能不能通过控制省份、文理等定类变量来得到可靠的结果呢?下面本文来说明,由于数据的样本数量有限,要做到这一点比较困难。

(二)放入控制变量能不能控制省份、文理差异

标准化难以消除省份、文理差异,那么在模型中控制这些变量是否能解决问题呢?理论上来说这是可行的,但是在“首都调查”中却难以做到这一点,因为样本量太小。在“首都调查”中来自某省的学生大概有多少呢?为了简明,假设调查样本中各个省份的学生数量一样多。如果不考虑拒访等情况,“首都调查”预计获得的总样本量为5100人。在平均水平下,数据中来自该省的人数应该有大约170名学生^①。当变量的方差较大而样本规模较小时,对总体参数的估计很可能无法得出显著的结果。不足200人的样本规模恐怕不算太大。实际情况是否如此呢?本文采用简单随机抽样从这个省所有考分不为0的考生中抽取170人形成模拟样本。在模型中控制省份、文理的变量,就相当于专门对某省、某一科类学生进行分析。模拟样本中有理工科考生116人,为了研究简明,本文只选取这一人数最多的类别进行分析。

研究教育不平等常用的模型是逻辑斯蒂回归,对于一个二分类的自变量,系数表示的是与参照群体相比,另一群体中事件发生的比数比的对数。比数指的是某个群体中事件发生的次数与不发生的次数之间的比值,而比数比就是对于两个群体比值做除法得出的商。在逻辑斯蒂回归中,如果因变量为“是否成功升学(进入某一层面的高校)”,性别自变量的参照群体是女生,那么自变量系数表示的含义就是“成功升学”这一事件在男生中发生的比数与在女生中发生的比数之间的商,再取对数。本文接下来计算系数时,都以女生为参照群体,因此如果系数为正,说明男生更占优势。

要获得全局的视野,可以采取“阈值依赖不平等”(李代,2017)的研究框架,计算每一种可能的分数线

^①实际上北京各高校在各省招生人数不一,在很多省份招生的规模低于全国平均水平。本文数据所来自的省份就是如此,所以170名学生是一个过高的估计。

下对应的系数大小。对原始的高考总体数据进行分析,比较男生与女生在升学上的差异,得到的结果是图2-c。对模拟抽样得到的116人数据进行同样的分析,得到的结果是图2-d。两相对比,二者存在着巨大的差异。统计模型也指出这一做法是不可靠的:用116人数据来估计分数是否超过三本线及以上所有假想分数线或者实际上是否被一本大学、二本及以上等层次大学录取的模型,自变量系数都是不显著的。当然,要从总体数据中随机抽取小样本的情况不只本文这一种,所得具体结果也可能与本文呈现出来的情况大相径庭。但这个例子背后的道理想必是清楚的,那就是样本量不足的时候难以得到统计显著的结果。

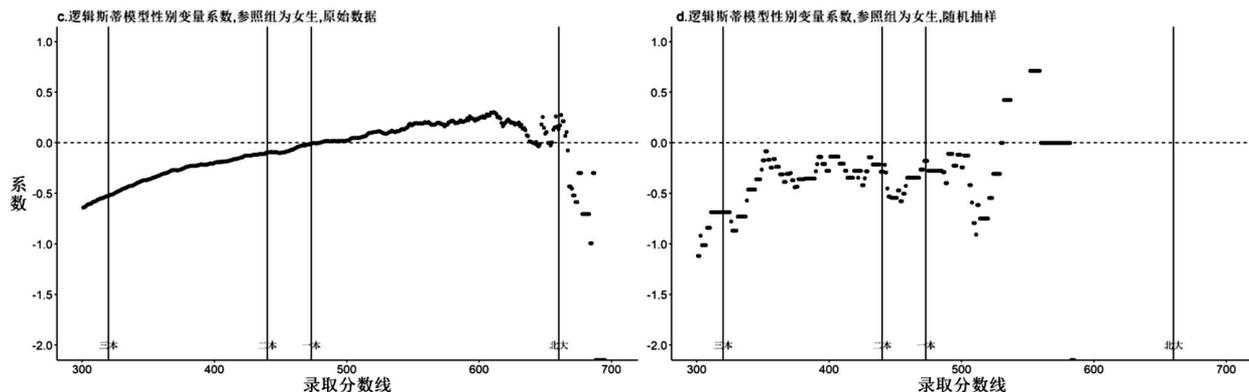


图2 随机抽样与原始数据计算得出的性别不平等分布

简单随机抽样是概率抽样的正统方法,因此问题没有出在抽样方法上。理论上讲,继续扩大样本量能够缓解这一问题,但是这也意味着更高的抽样成本。而且这里只观察了性别这一个变量的系数,如果还要考察更多变量,对样本规模的要求就更大了。所以,一般性抽样调查尽管在理论上能够很好地研究不平等,但受限于现实条件,结果未必尽如人意。

小结一下,由于升学是分省、分文理进行的,要解决生态谬误的问题就必须把这些因素带来的影响控制住。但是直接对分数进行标准化难以做到这一点,而把这些因素作为控制变量直接放入模型对样本量有较高的要求,一般性的抽样调查数据受限于现实条件,往往样本量不能满足模型的要求。

上面说的这两个问题,不只是对“首都调查”成立,一般而言对抽样调查都成立。下面本文进一步阐述“首都调查”作为目标地调查存在的两个问题:第一个问题属于一种可阐释性问题,第二个问题属于一种可推广性问题。在进行下面的分析时,所有的分析都立足于一省之内、只对理科进行,因此是剔除了生态谬误之后出现的问题。

三、作为目标地调查的问题

(一)可阐释性方面存在的问题

目标地调查的总体是一个特定的人群。例如“首都调查”当中,研究的总体是2006年和2008年入学、在北京54所公立高校就读的全日制本科学生。这个大学生群体是存在于日常经验当中的社会群体,先有了这个群体,后发生了大学经历。如果要研究大学经历对这些学生带来的影响,结果容易阐释。例如,通过研究得出“跟其他在北京上公立大学的学生相比,上北大、清华和人大的学生毕业后起薪平均高X元”这

类结论是容易理解的。

但是如果用同一个数据去研究哪些因素影响学生升学——也就是哪些因素影响学生进入这个“北京公立全日制大学生”群体,会带来阐释性上的问题。这是因为在升学发生的时刻,并不存在这个边界清晰的“即将到北京上公立大学”的群体。对于一个准备参加高考的学生来说,“跟所有其他将要到北京上公立大学的考生相比,女生比男生考上的可能性更大”是不容易理解的。这名学生既不确定自己属不属于“将要到北京上公立大学的考生”,其次不知道这个群体还有哪些其他成员。这种情况下,用目标地数据去进行定量研究,即便得到的结果是完全准确的,也难以与日常生活的经验进行沟通。

目标地数据中仅有成功者而无失败者,因此其总体与生源地数据不同,是升学过程的产物而非升学之时就存在于日常经验中的社会群体。它是为了统计研究的便利而限定出的“人造”总体,由此造成了对其进行研究的结论与日常经验沟通的困难。这个问题是一种可阐释性问题,它是目标地数据的核心问题。它并不是统计方法问题,而是研究意义问题。梁晨等(2012)、梁晨和董浩(2015)的研究,倡导使用新颖的学籍卡数据,回避了抽样的问题。对他们的研究,应星和刘云杉(2015)从理论上提出商榷,也对其材料与概念进行了质疑,关注的焦点之一便是数据研究的结果与历史研究的结果之间存在的逻辑断层。吴晓刚(2016)也指出,在解读数据时应该把成功升学的各人群比例与当时整个社会结构中各人群比例相对照才能得出对流动性的正确认识。本文认为即便是准确抽样甚至是总体数据,仅对目标地院校学生的构成比例进行描述统计,不但结果可能出现偏差,而且也难以阐释。

采用一般性抽样调查数据或者生源地数据就不存在这一问题。前者自不必说,后者构建的整体是“全省参加考试的学生”,对学生而言也是一个有意义的群体。当然不同的界定总体的方式,意味着可能会得出非常不同的结论——报名参加高考的人群结构本身已经可能和全省同期群的人口结构存在差异了,测量到的不平等系数也可能会发生变化,表示不同的含义,需要注意二者之间的差别。

(二)可推广性方面存在的问题

如果能把研究结论推广到日常经验当中存在的人群上,是不是就能解决可阐释性的问题了呢?例如,假如这一研究的结果可以推广到全部考生身上,可以推广到全国或者至少高等教育资源发达地区,是不是就能对结果做出合理的阐释了呢?要做到这一点,必须假设数据代表的总体与推广得到的总体之间是同质的,所以研究了前者也就研究了后者。但是“首都调查”数据在可推广性方面存在局限。下面假设“首都调查”获取的不是抽样调查数据而是总体数据,也就是北京全部公立高校全日制本科生的数据。这一总体中的性别不平等,与其他方式界定的总体之间存在多大的差异呢?

可推广性可以从纵向与横向两个维度进行考察。从纵向来看,可以考察成功升学的学生中存在的平等是否能推广到全体考生。对于目标地数据而言,由于总体中只有升学的成功者,因此只能进而对升学质量进行区别,从而形成因变量。在“阈值依赖不平等”的分析框架下,要比较的就是以分数超过某一分数线作为因变量,性别自变量的系数对于实际上成功升学的考生和全体考生而言是不是一样。从横向上看,可以考察成功升学的学生中存在的平等对于不同范围内学生是不是一样,例如北京大学生中的不平等与全国水平是否一样、与上海大学生中的不平等是否一样,北京公立大学生中的不平等与全部北京大学生的不平等是否一样等等。

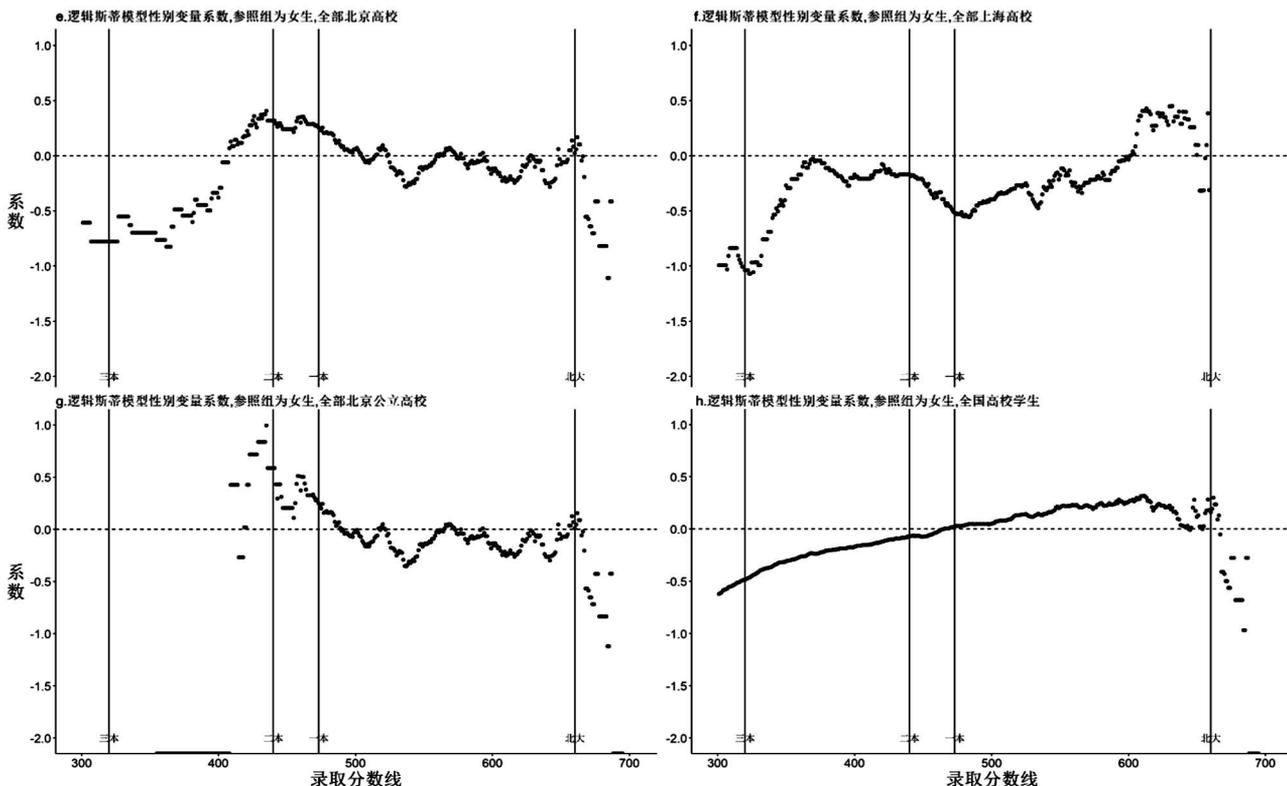


图3 不同总体的性别不平等

在纵向上,目标地数据的结果是否能推广到全体考生上去? 以比数比为例,对于群体X,事件*i*发生的次数是*a*,不发生的次数是*b*;对于群体Y,事件*i*发生的次数是*c*,不发生的次数是*d*。那么以Y为参照组,计算得出的比数比为 $\frac{a/b}{c/d}$ (1)

在研究升学时,事件*i*表示的可以是“被精英院校录取”,而群体X为男生、群体Y为女生。当把研究的总体从全体被大学录取的考生推广到全体考生上去时,新增那部分考生是那些参加考试但没有被大学录取的考生,因此在事件*i*不变的条件下,在两个人群中事件*i*发生的次数没有改变,而新增的群体X人数*m*和新增的群体Y人数*n*都归属到事件*i*不发生的情况里。这时,计算得到的比数比为 $\frac{a}{b+m} / \frac{c}{d+n}$ (2)

将(1)与(2)相除,得到 $\frac{d(b+m)}{b(d+n)}$ (3)

可以看到,要让(3)取值为1的条件是 $m/n = b/d$ 。换句话说,条件是全体考生中未被录取的男女性别比等于被录取的考生中未被精英大学录取的男女性别比。只要这个条件不能得到满足,缺乏失败者的目标地数据的结果就不能推广到全体考生上去。

与此同时,如果*m*和*n*的数值与*b*和*d*相比较小,那么即便上述条件不能得到满足,(3)的取值也与1相去不远。在本文采用的数据中,得到各类高校录取的理科考生占报考理科考生的比例达到82.6%,相应的*m*和*n*数值较小,比例也并不非常悬殊,因此系数值变化的幅度不大。将全国得到录取学生的情况(图3-h)与全体报考学生(图2-c)进行比较,后者整体向上有较小的移动,说明让(3)取值为1的条件并没有得到满足;但是变化幅度很小。严谨地讲,这种纵向上的推广大概率上存在偏误,但是如果推广带来的新增人数

不多,偏误的程度可能有限。

在横向上可以把对北京公立大学的考生情况推广到更广泛的群体,从而让结果更容易理解吗?横向推广与纵向本质的区别在于,在纵向推广中事件*i*发生的次数是不变的,而横向推广中事件*i*发生的次数也要改变,这就使得形式化的分析意义不大,不如直接看数据中呈现的结果。在这个维度上,又可以进行三组对比。

首先,将全部北京高校与全国高校做比,也就是说把所有被北京高校录取的学生(图3-e)和被全国高校录取的学生(3-h)进行比较。二者存在着明显的差异,说明北京高校学生的情况不能推广到全国高校中去。北京高校的招生情况与全国高校的整体情况相比有明显差异,这符合常识。

其次,将全部北京高校与全部上海高校做比,也就是说把所有被北京高校录取的学生(图3-e)和所有被上海高校录取的学生(图3-f)进行比较。二者存在着明显的差异,这说明北京高校学生的情况不能推论到上海高校的学生中去。北京和上海都是我国教育资源比较密集的地区,笼统地说看似有可比性,但数据分析显示的结果则是二者有较大差异。

最后,将全部北京高校与全部北京公立高校做比,也就是说把所有被北京高校录取的学生(图3-e)和所有北京公立高校录取的学生(图3-g)做比。“首都调查”对总体的限定是所有学生都来自北京的公立大学,因而是3-g图反映的情况。两相比较,可以看到不平等的分布在一本线以上基本保持相同,但在一本线之后就有了较大变化,尤其是在二本线之后。经过比较公立大学名单和高考数据中全部录取院校的名单,可以看到公立院校基本上就是在该省招生的全部一本和二本大学。两图对比而得的差异,源于学生的报考行为在一本线、一二本线之间和二本线以下有巨大差异。所以,这说明“首都调查”的研究成果只对公立高校成立,不能推广到全部北京高校中去。

纵向与横向两个维度上的可推广性问题,纵向上的问题属于目标地数据专有,而横向上的问题对于生源地数据也成立。例如,使用某一省生源地高考数据得出的研究结果恐怕也不能推广到其他省份乃至全国,因为各个省份存在较大差异。但是由于可阐释性上存在的问题,目标地数据的研究可能有更大的动机将其结果推广到广泛的人群上去,即便这并不可靠。对于生源地数据,研究结果限定于某一特定范围仍然是有意义的,因此研究者本不必将其具体结果推广到更大的范围。

上面描述的是总体数据的情况,还不涉及抽样方法问题。“首都调查”通过多阶段、分层、概率与规模成比例(PPS)抽样采集数据,然而该调查的抽样实施效果并不理想,因此样本数据对总体参数并不是无偏估计。严格来说,这一缺陷并不是目标地数据必然有的问题,而是这一调查自身的问题。

四、“首都调查”自身的抽样问题

“首都调查”采取了PPS的抽样方法,先在前面提到过的6层院校中分别抽取院校,再在抽到的15所院校中抽取学生。按照“首都调查”提供的被抽中的15所院校名单,笔者从高考完整数据中提取了全部相应高校的考生,并绘制了性别不平等的分布图(图4-i)。可以看到,在抽取院校这一阶段得到的结果不甚理想,不平等的分布与图3-g大相径庭。

这一偏误是不是因为没有加权呢?假如对数据进行合理的加权,是不是就能纠正偏误,对总体进行推断了呢?按照吴晓刚(2016)中的说明,笔者对数据进行了加权。根据北京公立高校各个大学在该省招

生的人数,可以算出6层高校每一层的总人数。对于北京大学、清华大学和中国人民大学这三层,各自的招生人数就是各层人数,因此无须处理。对于后三层高校,“首都调查”各抽出了6、2、4所高校。分别从这三层高校学生中有放回地抽取学生,使之达到这三层各自的总人数,也就得到了加权后的结果,见于图4-j。加权后与加权前相比改观不大,可见加权并不能很好地纠正这一阶段产生的偏误。这里进行的操作还是在理想情况下进行的,如果考虑到当“首都调查”进行加权的时候很可能没有考虑省份、文理的因素,对来自相同高校、不同省份或科类的学生会给予相同的权重,那么所加的权重很可能不会给减轻偏误带来帮助。

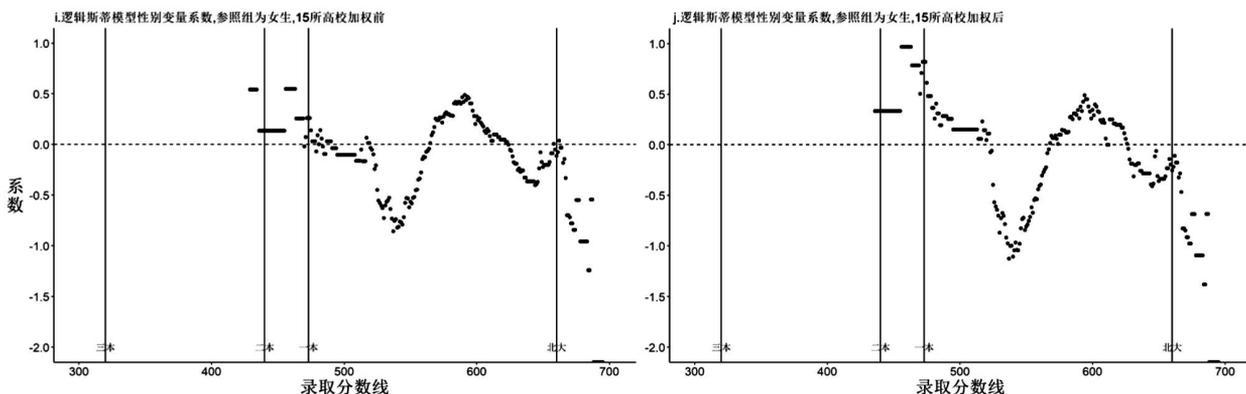


图4 分阶段抽样所得高校的性别不平等分布

这个偏误可能是由于多阶段抽样中抽取高校数量太少导致的^①。因为院校之间的异质性比较大,抽中了某个高校就会很大地影响结果的代表性。即便是同一层的高校之间本来也不是平行的关系,而在分数上存在着先后排名,在招生结构上存在较大差别。高分考生会尽量进入高排名高校,低分考生只能进入低排名高校。如果随机抽取高校的时候抽得过少,有可能会直接损失掉某个分段的部分学生,或者只抽到理工科较多而人文社科较少的院校,这样不论怎样对抽到的学生进行加权,都不可能弥补回失去的信息。理论上来说,通过增加在各层抽取高校的数量会有所帮助,但是也会相应增加调查实施的成本,因此未必具有可行性。

对于“首都调查”偏误的评估本身对基于这一数据进行的定量研究有重要的意义。基于这一数据进行的具体研究在多大程度上可能受到了偏误的影响?能否采取任何措施对这些偏误加以修正?是否能采用因果推断等研究框架缓解数据偏误带来的问题?这些都是使用这一数据的研究者应该澄清的问题。

本文的意图并不是批评这一调查在具体实施过程中出了差错。当实施抽样时,一旦具体步骤偏离了抽样方法在数学上要求的条件,其结果的可靠性也就失去了数学定律的支撑。在“首都调查”之外,也有不少调查因为受限于现实条件或偶然因素而采取了权宜之计的抽样方案。这些方案的代价有多大、其研究结果有多可靠,往往缺乏有效的手段来加以校验。这一困境是社会科学面临的普遍性问题,亟待认真的讨论。

① 实际上调查还在专业层次进行了阶段抽样,这对可推广性恐怕更加不利。

五、结论与讨论

上面的研究表明,要用“首都大学生成长追踪调查”研究高等教育机会的不平等,存在着若干缺陷和局限。有一些是抽样调查的一般性问题,有一些是作为目标地调查数据带来的问题,还有一些是调查实施中出现的问题。

一般性问题主要是生态谬误。因为高考录取是分省、分文理进行的,不对相关变量加以控制就难以避免生态谬误。要解决生态谬误可能有两个办法,一是标准化,二是控制变量。而至少在“首都调查”中,二者都不成功。使用其他一般性的社会调查——例如中国家庭追踪调查(CFPS)、中国综合社会调查(CGSS)、中国教育追踪调查(CEPS)等等时,也必须考虑这个问题。

作为目标地数据带来的问题包含两个层次:可阐释性与可推广性问题,最根本的是可阐释性问题。目标地调查构建的研究总体是升学的产物,是一个在升学发生时日常经验中尚不存在的群体,因此研究成果很难阐释成日常经验中有意义的命题。为了解决这个问题,研究者可能会把研究结论推广到其他在日常经验中存在的群体上去。从纵向上,把目标地数据的研究结果推广到包含了失败者的全部考生群体,从横向上,把研究结果推广到更广泛的人群上去。但是本文发现“首都调查”的结果不能推广至全体参加高考的考生、不能推广至全国、不能推广至教育资源同样密集上海地区、不能推广至三本院校。社会科学研究者自然可以明辨研究结果效力的限度,但普通的社会公众恐怕就不明白其中分寸何在。例如,2017年6月27日某网站以“北京大学生追踪调查:除了家庭收入,好中学更能影响上好大学”^①为题对“首都调查”和吴晓刚(2016)的研究成果进行了报道。报道把研究成果转述为“家庭社会经济地位和居住地对于进入不同层次的高校仍然有直接的影响,但对进入什么层次的大学作用更加明显是,是否来自重点高中,以及是否获得高考特殊政策的照顾”。这种措辞方式中总体似乎是全部考生,没有考虑实际上这一调查研究总体的范围仅为北京公立高校的学生,不能说有一定误导性。

目标地数据在可推广性方面存在的问题,本质是不是选择性偏误的问题呢?二者之间存在联系,但是又有必要进行区别。由于种种原因,考生在选择去哪个省份的高校就读时不是随机的,因此,北京的情况跟上海存在差别,更与全国情况存在差异。如果研究声明的总体是全国高校学生,而在抽样时只对北京市高校的学生进行抽样,那显然存在选择性偏误的问题。但是如果研究清楚地声明自己的研究总体是北京市高校的学生,也对这个总体进行了正确的抽样,所得的结果能够正确地估计该总体的参数,也就不能说存在选择性偏误的问题。本文讨论的偏差出现在将某一局部的情况推广到更广泛的群体上的时候,对这一问题用“可推广性”来指认更不容易带来混淆。

如果把目标地的数据收集的范围扩大到全国大学生,是不是就解决了可阐释性和可推广性的问题呢?首先,学者们已经指出,这种情况下数据中仍然只有成功者而无失败者,因此存在局限(吴晓刚,2016)。其次,“将要上大学”这个群体尽管比“将要去北京上大学”听上去更符合日常经验,但仍然不完全是一个边界清晰的社会群体。尤其是对于中低分数段的考生而言,能否上大学、是否上大学都是值得认真权衡的问题,包含着不确定性。总之,将目标地数据的范围扩大,对问题有所缓解,但并没有

^① 新闻原文可参见:http://www.thepaper.cn/newsDetail_forward_1718238.

解决。

“首都调查”自身还存在抽样实施中的问题。它在抽取高校的阶段可能产生了系统性偏误,让样本不能提供对总体参数的无偏估计。这一情况的发生固然非常不幸,但恐怕也难以事先避免。如果在构建抽样框的时候就能掌握全部学生的信息,那就不需要进行分阶段抽样;进行了分阶段抽样就意味着出于某些现实的原因而可能无法掌握全部学生的信息,那也就不具备很好的条件来判断多阶段抽样的代价如何。

本文提出的问题如何得到有效的解决?在理想条件下,如果能够把生源地数据和目标地数据打通,就能更好地对双方进行校验。在研究高等教育机会的问题上,中国具有得天独厚的优势,因为高校录取以标准化考试为主要依据、招生过程统一进行而且规则清晰,因此相关数据如能提供给研究者、甚至与各大学管理的学籍数据、民政部门拥有的户籍数据联通,必能对研究高等教育机会的问题带来裨益。这当然不是一个新想法,只是要实现它并不取决于研究者。在现实条件下,采集目标地数据仍然将是对研究者而言最有可行性的调查方案,因为它成本更低。如果研究者能明辨其研究结果的限度,审慎地评价其可阐释性、可推广性的程度,评估结果偏误的程度,并尽可能地使研究趋于严谨,那么即便不能得到完全准确的结果,也能给学术研究带来贡献。如果能进一步地提出新的研究方法,克服上述种种局限,对学术研究的贡献就更大了。

参考文献:

- 陈伟、乌尼日其其格,2016,《职业教育与普通高中教育收入回报之差异》,《社会》第2期。
- 李代,2017,《阈值依赖的教育扩张与教育机会不平等——以a省某年高考数据为例》,《社会学研究》第3期。
- 李骏,2016a,《城乡出身与累积优势:对高学历劳动者的一项追踪研究》,《社会学研究》第2期。
- ,2016b,《中国高学历劳动者的教育匹配与收入回报》,《社会》第3期。
- 李路路主编,2013,《中国大学生成长报告2012》,北京:中国人民大学出版社。
- 李忠路,2016,《家庭背景、学业表现与研究生教育机会获得》,《社会》第3期。
- 李忠路、邱泽奇,2016,《家庭背景如何影响儿童学业成就?——义务教育阶段家庭社会经济地位影响差异分析》,《社会学研究》第4期。
- 梁晨、董浩,2015,《必要与如何:基于历史资料的量化数据库构建与分析——以大学生学籍卡片资料为中心的讨论》,《社会》第2期。
- 梁晨、董浩、任韵竹、李中清,2017,《江山代有才人出——中国教育精英的来源与转变1865—2014》,《社会学研究》第3期。
- 卡尔·波普尔,1987,《客观知识——一个进化论的研究》,舒炜光等译,上海:上海译文出版社。
- 吴晓刚,2016,《中国当代的高等教育、精英形成与社会分层》,《社会》第3期。
- 谢桂华、张阳阳,2016,《向党靠拢被党接纳:大学生入党问题研究》,《社会》第3期。
- 许多多,2017,《大学如何改变寒门学子命运:家庭贫困、非认知能力和初职收入》,《社会》第4期。
- 杨中超,2016,《教育扩招促进了代际流动?》,《社会》第6期。
- 叶晓阳、丁延庆,2015,《扩张的中国高等教育:教育质量与社会分层》,《社会》第3期。
- 张春泥、谢宇,2017,《入学年龄限制真的造成了“七八月陷阱”吗?——兼评刘德寰、李雪莲〈“七八月”的孩子们〉》,《社会学研究》第1期。
- 朱斌,2018,《文化再生产还是文化流动?——中国大学生的教育成就获得不平等研究》,《社会学研究》第1期。
- Simpson, E. H. 1951. "The Interpretation of Interaction in Contingency Tables" *Journal of the Royal Statistical Society, Series B.* 13: 238-241.

Limitations of Using Target–Location Survey Data in Understanding Educational Inequality: The Case of Beijing College Students Panel Survey

LI Dai

Abstract: Educational inequality is an important field in the study of social stratification. Some quantitative projects may use target–location data, which are data gathered from college students, to study educational inequality. For example, Beijing College Students Panel Survey (BCSPS) is a target–location data. This article compares simulated data generated in the same way as BCSPS did with a complete data of National College Entrance Examination from a province in the threshold–dependent inequality framework. Results show that the estimates derived from such data are faced with difficulties in interpretability and generalizability. BCSPS particularly suffers bias from its sampling procedures. These limitations may lead to misleading conclusions.

Key Words: Educational Inequality; National College Entrance Examination; Social Stratification; Social Mobility

(责任编辑:唐丽娜)